

NAVAL POSTGRADUATE SCHOOL

Monterey, California



THESIS

**IMPROVED QUESTIONNAIRE DATA HANDLING AND
AUTOMATED ANALYSIS CAPABILITIES FOR THE
MV-22 OPERATIONAL EVALUATION**

by

William K. Hershberger

September 1999

Thesis Advisor:
Second Reader:

Lyn R. Whitaker
Thomas. H. Hoivik

Approved for public release; distribution is unlimited.

19991126 070

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

1. AGENCY USE ONLY (Leave Blank)		2. REPORT DATE September 1999	3. REPORT TYPE AND DATES COVERED Master's Thesis
4. TITLE AND SUBTITLE IMPROVED QUESTIONNAIRE DATA HANDLING AND AUTOMATED ANALYSIS CAPABILITIES FOR THE MV-22 OPERATIONAL EVALUATION		5. FUNDING NUMBERS	
6. AUTHOR(S) Hershberger, William K.		8. PERFORMING ORGANIZATION REPORT NUMBER	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Naval Postgraduate School Monterey, CA 93943-5000			
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)		10. SPONSORING / MONITORING AGENCY REPORT NUMBER	
11. SUPPLEMENTARY NOTES The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government.			
12a. DISTRIBUTION / AVAILABILITY STATEMENT Approved for public release; distribution is unlimited.		12b. DISTRIBUTION CODE	
13. ABSTRACT (Maximum 200 words) This thesis investigates the method by which the V-22 Multiservice Operational Test Team (MOTT) collects and analyzes subjective operational test data, and suggests improvements to the tools currently in use. The MOTT uses a Microsoft Access custom application to generate and administer questionnaires and report summary results of the data collected. Using Visual Basic for Applications to implement Automation techniques, specific program improvements are created and implemented that use basic summary statistics to assist in handling the problems of excessive questionnaire lengths and user-friendly summary report generation. A trend analysis is conducted on data from a previous operational assessment in terms of MOTT pilot demographics to highlight any response biases exhibited by subsets of participants. Analysis techniques are demonstrated and specific findings are discussed in terms of potential to affect future test results. In addition, cluster analysis is used to investigate an improved structure for the generic maintenance questionnaire used in a previous operational assessment. Suggestions are put forth for aggregating questions and creating a more efficient maintenance questionnaire.			
14. SUBJECT TERMS Test and Evaluation, Operational Testing, Survey Analysis, Computing and Software, Questionnaires, V-22 Osprey, VBA, OLE Automation		15. NUMBER OF PAGES 121	
		16. PRICE CODE	
17. SECURITY CLASSIFICATION OF REPORT Unclassified	18. SECURITY CLASSIFICATION OF THIS PAGE Unclassified	19. SECURITY CLASSIFICATION OF ABSTRACT Unclassified	20. LIMITATION OF ABSTRACT UL

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. Z39-18

Approved for public release; distribution is unlimited.

**IMPROVED QUESTIONNAIRE DATA HANDLING AND AUTOMATED
ANALYSIS CAPABILITIES FOR THE MV-22 OPERATIONAL EVALUATION**

William K. Hershberger
Major, United States Marine Corps
B.A.S., University of Washington, 1986

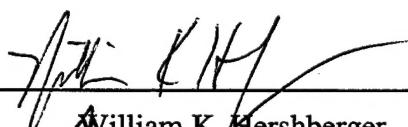
Submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE IN OPERATIONS RESEARCH

from the

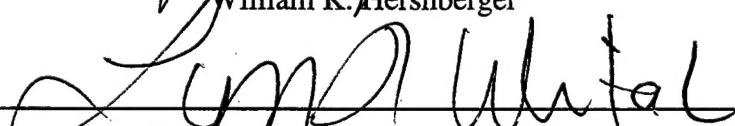
NAVAL POSTGRADUATE SCHOOL
September 1999

Author:

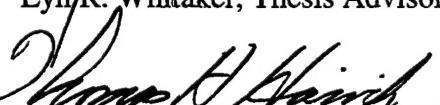


William K. Hershberger

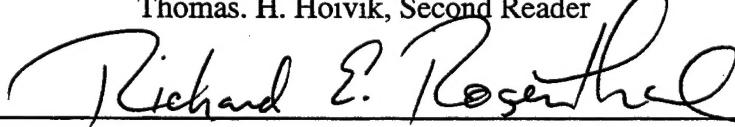
Approved by:



Lyn R. Whitaker, Thesis Advisor



Thomas. H. Hoivik, Second Reader



Richard E. Rosenthal, Chairman
Department of Operations Research

ABSTRACT

This thesis investigates the method by which the V-22 Multiservice Operational Test Team (MOTT) collects and analyzes subjective operational test data, and suggests improvements to the tools currently in use. The MOTT uses a Microsoft Access custom application to generate and administer questionnaires and report summary results of the data collected. Using Visual Basic for Applications to implement Automation techniques, specific program improvements are created and implemented that use basic summary statistics to assist in handling the problems of excessive questionnaire lengths and user-friendly summary report generation. A trend analysis is conducted on data from a previous operational assessment in terms of MOTT pilot demographics to highlight any response biases exhibited by subsets of participants. Analysis techniques are demonstrated and specific findings are discussed in terms of potential to affect future test results. In addition, cluster analysis is used to investigate an improved structure for the generic maintenance questionnaire used in a previous operational assessment. Suggestions are put forth for aggregating questions and creating a more efficient maintenance questionnaire.

DISCLAIMER

The reader is cautioned that computer programs developed in this research may not have been exercised for all cases of interest. While every effort has been made, within the time available, to ensure that the programs are free of computational and logic errors, they cannot be considered validated. Any application of these programs without additional verification is at the risk of the user.

TABLE OF CONTENTS

I.	INTRODUCTION.....	1
A.	BACKGROUND.....	2
B.	PROBLEM DEFINITION	4
1.	Length of Questionnaires	4
2.	Inconsistencies in Responses.....	5
3.	Presentation of Results	5
C.	THESIS OBJECTIVE	6
1.	Reduce Questionnaire Length	6
2.	Improve Reliability of Data.....	6
3.	Enhance OTD Interpretation of Results	6
D.	THESIS SCOPE AND ORGANIZATION.....	7
II.	VQT AND THE V-22 TEST PROGRAM.....	9
A.	V-22 STATUS.....	9
B.	THE ROLE OF VQT	14
1.	Questionnaire Creation.....	16
2.	Questionnaire Administration	19
3.	Summary Reporting	21
III.	PRELIMINARY ISSUES	25
A.	PROGRAMMING ENVIRONMENT	25
1.	Software.....	25
2.	OLE Automation	26
B.	DATABASE NORMALIZATION	27
C.	MEASUREMENTS	30
IV.	PROGRAMMING METHODOLOGY	37
A.	CRITERION-BASED SCREENING.....	37
B.	REPORTING.....	42

V. OT-IID DATA ANALYSIS	51
A. PILOT DEMOGRAPHICS	51
1. Data	51
2. Method of Analysis	53
3. Results	63
B. MAINTENANCE QUESTIONNAIRE	66
1. Data	66
2. Method of Analysis	67
3. Results	74
VI. CONCLUSIONS	77
A. DISCUSSION	77
1. Questionnaire Length	77
2. Reliability of Data	78
3. OTD Interpretation of Test Results	79
B. RECOMMENDATIONS	80
1. Pilot Data.....	80
2. Maintenance Questionnaire	81
C. FURTHER RESEARCH OPPORTUNITIES	81
APPENDIX A. VQT DATABASE RELATIONSHIP DISPLAYS.....	83
APPENDIX B. VQT QUESTIONNAIRE SCREENING FUNCTION	85
APPENDIX C. AUTOMATION ROUTINE FOR CHART CREATION.....	89
LIST OF REFERENCES	95
INITIAL DISTRIBUTION LIST	97

LIST OF FIGURES

Figure 1.	Questionnaire Editing Form.....	17
Figure 2.	Mission Area Selection Form.....	18
Figure 3.	Questionnaire Administration Form.....	20
Figure 4.	Example Page of VQT Summary Report.....	22
Figure 5.	Histograms of Example Questionnaire Data Distributions	33
Figure 6.	Revised Mission Area Selection Form.....	39
Figure 7.	Screening Criteria Selection Form	40
Figure 8.	Example of Screening Results Form.....	41
Figure 9.	Revised Reports Menu Form	45
Figure 10.	Typical Summary Chart	47
Figure 11.	Sample Page of Summary Charts.....	49
Figure 12.	Pilot Histograms by Response Value.....	54
Figure 13.	Pilot Response Trends by Category	56
Figure 14.	Average Responses by Pilot for All Categories.....	57
Figure 15.	Average Responses by Aircraft Community.....	58
Figure 16.	Average Responses by Pilot Rank	59
Figure 17.	Average Response by Flight Hours.....	60
Figure 18.	Average Response by WTI Graduate.....	61
Figure 19.	Average Response by Service	61
Figure 20.	Cluster Analysis Plots on Complete Maintenance Data Set	69
Figure 21.	Maintenance Questionnaire Response Patterns.....	71
Figure 22.	Cluster Analysis Results on Abbreviated Data Set	72
Figure 23.	Cluster Analysis Results Using “Manhattan” Distance	73
Figure 24.	Original VQT Database Relationships.....	83
Figure 25.	Revised (Normalized) VQT Database Relationships.....	84

LIST OF TABLES

Table 1.	USMC OT-IID COIs.....	11
Table 2.	USAF OT-IID COIs	11
Table 3.	Assault Support COI/MOE/MOP Hierarchy.....	13
Table 4.	Comparison of Relative Efficiency Between VQT Versions.....	30
Table 5.	Questionnaire Response Definitions	31
Table 6.	80/50 Versus Mean & Standard Deviation	35
Table 7.	80/50 Versus Median & Interquartile Range	35
Table 8.	Mean & Standard Deviation Versus Median & Interquartile Range	35
Table 9.	MOTT Pilot Demographics.....	51
Table 10.	Mission Area Category Assignments	52
Table 11.	Pivot Table Report of Category Four Pilot Responses	63
Table 12.	Original Maintenance Questionnaire	66
Table 13.	Numbering Scheme for Abbreviated Questionnaire	71
Table 14.	Reworded Maintenance Questionnaire	74

LIST OF SYMBOLS, ACRONYMS AND/OR ABBREVIATIONS

ACAT	Acquisition Category
AFOTEC	Air Force Operational Test and Evaluation Center
ART	Automated Rating Tool
CH-46E	“Sea Knight”, Marine Corps medium lift assault helicopter
CH-53D	“Sea Stallion”, Marine Corps medium lift transport helicopter
CH-53E	“Super Stallion”, Marine Corps heavy lift transport helicopter
COI	Critical Operational Issue
DT	Developmental Test
MH-53J	“Pave Low”, Air Force long-range special operations helicopter
MOE	Measure of Effectiveness
MOP	Measure of Performance
MOS	Measure of Suitability
MOTT	Multi-service Operational Test Team
OLE	Object Linking and Embedding
OPEVAL	Operational Evaluation
ORD	Operational Requirements Document
OT	Operational Testing
OTD	Operational Test Director
OT&E	Operational Test and Evaluation
OT-IID	Operational Assessment of September-October 1998
OT-IIE	OPEVAL, scheduled for October 1999 – April 2000
TEMP	Test and Evaluation Master Plan
USAF	United States Air Force
USMC	United States Marine Corps
VBA	Visual Basic for Applications
VQT	V-22 Questionnaire Tool
WTI	Weapons and Tactics Instructor Course

EXECUTIVE SUMMARY

The V-22 Osprey is currently in the midst of operational test and evaluation (OT&E) and is scheduled to begin its Operational Evaluation (OPEVAL) in October 1999. This seven-month evaluation will involve the collection, analysis, and summary of vast quantities of objective and subjective data. The primary means by which the V-22 Multi-service Operational Test Team (MOTT) collects subjective test data is through the administration of questionnaires, which elicit numerical ratings and commentary from the respondents. In order to facilitate this task, the MOTT analysis section has created a custom database program which they named the "V-22 Questionnaire Tool", or VQT.

VQT has successfully automated parts of the questionnaire administration process. But the program has several shortcomings in terms of questionnaire construction methods, data analysis, and presentation of results. The questionnaires that are generated to collect data from a test event have typically been excessive in length, which brings into question the reliability of the data collected from these questionnaires. The reports that are generated by the program, although comprehensive and not difficult to interpret, are purely textual and do not effectively summarize the data. No other automated analysis is provided. And in past operational assessments, there have been instances of questions that have elicited inconsistent response patterns that cause difficulties in accurately summarizing the results of the test. This thesis proposes methods to alleviate these shortcomings in the current data collection method.

Two distinct efforts are made in this thesis; programming and analysis. The programming aspect focuses on two objectives; to provide an additional screening

capability for the questionnaire creation process that will help make shorter questionnaires, and to provide the ability to create a series of charts that effectively summarize the data collected up to that time. These functions are developed using Visual Basic for Applications (VBA) within Access 97. The analysis focuses on two data sets from the questionnaires administered during the last operational assessment; V-22 pilot responses, and responses to the standard maintenance questionnaire that was administered for all maintenance actions. A trend analysis is conducted on the pilot data to identify any response patterns, biases, or inconsistencies that have the potential to effect test results. The maintenance data is examined using cluster analysis, to explore possible ways to reduce the length of the maintenance questionnaire.

A screening function is created by which the user can provide criteria that VQT uses to screen questions prior to including them on a questionnaire being developed. The user provides a threshold mean, standard deviation, and sample size. The program then uses these criteria to determine the adequacy of responses already provided for each question being considered for inclusion on a questionnaire. Questions that meet or exceed the threshold are considered as being sufficiently addressed, and are not included in the questionnaire. Consistent use of this capability in the future should result in progressively shorter questionnaires and a necessarily increased emphasis on the questions that require more attention and evaluative effort.

A charting function is developed for VQT that provides a graphical summary of numerical questionnaire responses. Using VBA to control Excel functions from within Access, a control is provided that instructs VQT to export the desired data set to Excel,

sort it appropriately, and create a series of charts that portray the mean and sample size for each question in the subset designated by the user. This provides a data summary that previously was not available within VQT. It is available at any time, facilitates an overall understanding of test results, and should assist the analysts and operational test directors (OTD) in questionnaire development, event planning, and writing the final test report.

The trend analysis of the pilot data demonstrates a technique using basic summary statistics that can be used to identify trends and biases in the response patterns and inconsistencies in the data. Several trends are identified in the OT-IID pilot data that warrant further investigation by the MOTT analysts and OTDs prior to OPEVAL. Using such a method during future test periods should facilitate identification of potential problems with the collection effort, problems that may bring into question the reliability of the data if appropriate corrective measures are not taken.

With the analysis of the maintenance questionnaire, a formal statistical method is demonstrated by which a repeatedly administered questionnaire may be evaluated in terms of its potential to be made smaller. Incorporating the results of a cluster analysis on the OT-IID maintenance questionnaire results in a 25% shorter example questionnaire that should provide nearly, if not exactly, the same information as the original. Using this analytical method, the analysts could use previously collected data to make the ongoing data collection process more optimal by administering a smaller, more efficiently constructed maintenance questionnaire.

ACKNOWLEDGMENT

The author would like to acknowledge the generous and enthusiastic assistance of LtCol. Mark Cwick, USMC, without whose foresight and unselfish dedication to his work this opportunity for research would probably not have existed. His memory will live on in those who had the pleasure of working with and befriending him, and his positive attitude and devotion to duty should serve as inspiration to us all.

I. INTRODUCTION

The V-22 Osprey tiltrotor aircraft is here. On 14 May 1999, the Marine Corps accepted from Bell Helicopter Textron the first of four low-rate initial production (LRIP) MV-22s to be delivered in 1999 [Ref. 1:p. 6]. As the long-awaited medium lift replacement for the Marine Corps' aging fleet of transport helicopters, the Osprey is in the midst of operational test and evaluation (OT&E) and is scheduled to begin its Operational Evaluation (OPEVAL) in October 1999. This approximately seven-month critical evaluation will involve the collection, analysis, and summary of vast quantities of objective and subjective data. The primary means by which the V-22 Multi-service Operational Test Team (MOTT) collects subjective test data is through the administration of questionnaires. In order to facilitate this task, members of the MOTT analysis section have created a database program which they named the "V-22 Questionnaire Tool", or VQT.

VQT has successfully automated parts of the questionnaire administration process. Collection, management, and rapid accessibility of large quantities of data by the analysts have been considerably enhanced. But VQT has several shortcomings in terms of questionnaire construction methods, data analysis, and presentation of results. To facilitate accurate and timely evaluation of operational test data while ensuring confidence in the reliability of the data, these shortcomings must be resolved prior to the initiation of OPEVAL. The purpose of this thesis is to address these issues by providing additional database capabilities and statistical methods by which to improve the process.

A. BACKGROUND

The V-22 is scheduled to begin replacing the Marines' aging fleet of CH-46 and CH-53D transport helicopters in late 1999. As the first production tiltrotor aircraft, the Osprey possesses payload, speed, and range capabilities far beyond those of current Marine helicopters [Ref. 2]. In conjunction with developmental testing, the Osprey has been undergoing a series of initial operational assessments over the last five years to judge the potential effectiveness and suitability of the airframe in an operational environment. The MOTT, a combined Marine Corps (USMC) and Air Force (USAF) team, operates out of the Patuxent River Naval Air Station and is conducting the operational testing. Members of the MOTT include Air Force and Marine Corps pilots from several aircraft communities, aircrew, maintainers, analysts, and administrative support personnel.

The questionnaires used to collect subjective test data are generated by the analysts for each event during the test period and filled out by members of the MOTT and other personnel participating in the event. Rapidly creating effective and complete post-test event questionnaires that are appropriate to a mission profile that may be constantly shifting is a challenging task. In addition, the evaluation of the Critical Operational Issues (COIs) enumerated in the test plan is dependent on the responses to several hundred unique questions, each of which addresses a specific system capability or mission area. The responses to all of these questions must be consolidated, validated, and analyzed. VQT is a Microsoft Access database that was created by the MOTT analysis section specifically to support and automate this process. [Ref. 3]

In order to automate the generation of questionnaires, the database (VQT) is populated with questions that relate to the specific Measures of Performance (MOP), Measures of Effectiveness (MOE), Measures of Suitability (MOS), or Critical Operational Issues (COI) enumerated in the Test Plan. Each question consists of two parts. The first is a rating of the system or function in question on a scale of six response alternatives, scored for analysis from one (worst) to six (best), and based upon the respondent's professional opinion. The second is the comment block, where the respondent is asked to provide amplifying information.

Prior to each scheduled test event, an analyst creates a questionnaire that is specific to that event and contains questions that are pertinent only to that event. Since each database record containing a question also contains a reference to the specific mission area(s) to which that question is pertinent, by having knowledge of the planned objectives for a test flight the analyst can quickly build a questionnaire based on the relevant mission areas and include all questions that are pertinent to that event.

Because it is resident on the MOTT computer network, VQT also automates the administration of questionnaires and consolidation of data to a large extent. When a test pilot returns from a flight, he fills out the appropriate questionnaire for his particular event. Normally this is accomplished by sitting at an office computer, logging in, and completing the questionnaire through the VQT interface designed for that purpose. Thus the data is entered directly into the database.

Currently VQT provides limited analysis and summary of the data. Results for each question are provided in the form of a summary report that lists the mean rating for

all responses, the total number of responses, the number of respondents who answered with each particular rating (1-6), and a listing of all commentary responses. All information is provided textually. Upon completion of a test period, each Operational Test Director (OTD) is provided with the VQT report(s) pertinent to the COI(s) for which he is responsible. He then uses that information to facilitate his write-up of the appropriate portion of the Test Report. [Ref. 3]

B. PROBLEM DEFINITION

While VQT has considerably enhanced the MOTT analysis section's ability to dynamically create and administer questionnaires, collect and consolidate data, and report summary results, there are several aspects of the program that warrant further examination.

1. Length of Questionnaires

Because mission profiles vary from day to day, the requirement for the pilot or other participant to evaluate various mission areas may vary drastically from one flight to the next. He may face a post-flight questionnaire consisting of 20 questions one day, while after the next day's event he faces 80 questions [Ref. 3]. This is, in part, an unfortunate consequence of automating the creation of questionnaires, since the questionnaire construction function of VQT queries the database for all questions pertinent to a given mission area. Therefore, to cover a flight composed of tasks from multiple mission areas (for example, a shipboard launch to a night troop insert followed by an aerial refuel) the length of the resulting questionnaire can be daunting. The

reliability of data collected from a questionnaire of 80 or more subjective questions must be called into question.

2. Inconsistencies in Responses

A number of questions have elicited numerical responses that are not consistent with the associated commentary and subsequent write-up in the Test Report. Two potential sources for the inconsistencies are immediately apparent. The first is the differing operational backgrounds and therefore perspectives of the MOTT pilots. In some evaluated areas that individual's perspective could foreseeably skew his perception of an item or capability [Ref. 3]. A second possible source for inconsistent responses is the questions themselves. Without diminishing the efforts of those who created the database, there do exist a number of confusing, vague, and otherwise poorly worded questions within VQT, which may elicit various responses depending upon an individual's interpretation of the question.

3. Presentation of Results

The reports that VQT generates, although complete in terms of data presentation, are entirely textual and are difficult to summarize and comprehend. Summary information and a complete list of comments are provided for every question pertinent to a specific COI, and because of the wide scope of some COIs and the resulting quantity of data required for evaluation, the size of the report generated is overwhelming. The non-statistically-trained OTDs are required to summarize this data and document the results in the final Test Report, a difficult task that could be aided by a more effective manner of summary presentation.

C. THESIS OBJECTIVE

The objective of this thesis is to enhance the capabilities of VQT by adding additional functionality to the program that incorporates basic analytical tools. An analysis of data gathered from a previous operational assessment (OT-IID) will also attempt to uncover any biases or trends that may be manifesting themselves in the response data. Specific objectives include the following:

1. Reduce Questionnaire Length

As it directly puts the validity of questionnaire data in doubt, the issue of questionnaire length is of primary importance to the utility of VQT as an analytical tool. This includes not only pilot questionnaires, but questionnaires designed for maintenance personnel and other event participants.

2. Improve Reliability of Data

Certain personal or community factors may predispose an individual to demonstrate bias with regards to certain mission areas or capabilities. Knowledge of these tendencies or biases will allow the MOTT analysts to control for them in an appropriate manner.

3. Enhance OTD Interpretation of Results

Pictorial or graphic representation of the results in a form tailored for non-statistically trained personnel will assist in accurate and complete understanding of the data collected. Incorporating this capability in a dynamic manner will also assist the MOTT in planning future flight events and allow them to monitor the daily progress of the operational test.

D. THESIS SCOPE AND ORGANIZATION

This thesis focuses on two primary areas of interest; programming and analysis. The programming is accomplished with Visual Basic for Applications (VBA), and remains within the confines of the original VQT program as it stood at the completion of the most recent operational assessment, OT-IID. Because the V-22 has already progressed through several operational assessments and data collection methods are fairly well established, the focus is on working within the established program to enhance current methods.

The questionnaire data collected during OT-IID is used as a basis for the analysis portion of the thesis. The analysis conducted is limited to the identification of trends within the pilot data, and on identifying similarities among responses to certain groups of questions on the maintenance questionnaire. Although many other individuals besides pilots and maintainers contributed to the OT-IID database, the focus is on the data collected from these two groups due to their prominent role throughout the test period and their potential to influence test results.

The section organization of this thesis will proceed along the following lines. Chapter II will present the status of the V-22 program, the background and construction of the VQT database, and how VQT fits into the V-22 testing process. Chapter III will discuss the preliminary assessment of the database and data, general program implementation concerns and limitations, and exploratory work required prior to design implementation and analysis. Chapter IV will examine the programming methods and processes that are used to create the additional database functions. Chapter V will

discuss the methods used to analyze the OT-IID pilot data and maintenance questionnaires. Finally, Chapter VI concludes the report with a summary of results and findings, a list of specific recommendations based on the analyses conducted, and suggested topics for further research.

II. VQT AND THE V-22 TEST PROGRAM

A. V-22 STATUS

A program that was at one time officially cancelled by the Secretary of Defense, the V-22 is now well on its way to becoming the primary assault support aircraft in the Marine Corps inventory. It is a key aspect of the Operational Maneuver From The Sea (OMFTS) doctrine of the United States Navy and Marine Corps, and brings a vastly increased power projection capability to the Marine Air-Ground Task Force (MAGTF). An ACAT-1C program anticipated to have a total program cost of \$37 billion, the V-22 is pending a successful Milestone-3 decision, approval for the first full production lot, contingent upon successful completion of OPEVAL beginning in October 1999 [Ref. 4].

Because of the inherent flexibility and performance envelope which is greatly expanded from that of a traditional assault helicopter, the V-22 has potential utility in other mission areas besides assault support, a fact which has not passed by the Air Force Special Operations Command. The Air Force has committed to assisting the Marine Corps in the development and evaluation of this aircraft, which has resulted in the formation of the MOTT and the direct participation by Air Force pilots and other personnel in the operational testing of the V-22. The unique mission requirements of the special operations community have resulted in an expanded set of operational test criteria which must be met, although not all criteria apply to the MV-22. The CV-22 will be the special operations version of the Osprey, the first one of which is not scheduled to be delivered to the Air Force until the year 2003 [Ref. 5]. But because a majority of the

operational requirements enumerated in the Operational Requirements Document (ORD) are common to both the MV-22 and the CV-22, the Air Force is taking an active role in the current operational testing of the MV-22. When the MV-22 (presumably) receives approval to continue on into the full-rate production phase, the Air Force will continue on with the evaluation of the special operations version of the aircraft.

As defined by the Defense Systems Management College, the purpose of operational testing (OT) is to evaluate the test item in terms of operational effectiveness, suitability, and need for further development, to provide information on organization and personnel requirements, to develop doctrine and tactics, and to verify operating instructions, documentation, and other publications [Ref. 6:p. 11-1]. Whereas developmental testing (DT) evaluates more technical criteria under controlled conditions, and can directly measure parameters under conditions which can be repeated, operational testing puts the item in a more realistic environment and examines it in terms of its contribution to, and ability to work within, the fleet operating forces. The test item is evaluated in terms of operational requirements, which may be far less well defined than the technical criteria examined in DT.

Because of the wide range of potential missions, environments, and threat scenarios in which it will be required to perform, the V-22's list of operational requirements is long and varied. The requirements enumerated in the Test and Evaluation Master Plan (TEMP) that are relevant to this thesis are the Critical Operational Issues (COI). As defined by the Defense Systems Management College, Critical Operational Issues are:

Operational effectiveness and operational suitability issues (not parameters, objectives, or thresholds) that must be examined in operational test and evaluation (OT&E) to determine the system's capability to perform its mission. A COI is normally phrased as a question that must be answered in order to properly evaluate operational effectiveness (e.g., "Will the system detect the threat in a combat environment at adequate range to allow successful engagement?") or operational suitability (e.g., "Will the system be safe to operate in a combat environment?"). [Ref. 6:p. B-6]

Among the COIs, there are six effectiveness issues (four USMC and two USAF) and fourteen suitability issues (twelve USMC and two USAF) enumerated in the TEMP that must be evaluated by the test team during OPEVAL. These COIs are given in Tables 1 and 2. [Ref. 7]

Operational Effectiveness	Tests
Assault Support	E-1
Self-Deployment	E-3
MV-22 Survivability	E-4
Tactics	E-6

Operational Suitability	Tests
Reliability	S-1
Maintainability	S-2
Availability	S-3
Logistics Supportability	S-4
Compatibility	S-5
Interoperability	S-6
Training	S-7
Human Factors	S-8
Safety	S-9
Documentation	S-10
Diagnostics	S-11
Software	S-12

Table 1. USMC OT-IID COIs

Operational Effectiveness	Tests
Special Operations Forces (SOF)	E-2
Self-Deployment	E-3
CV-22 Survivability	E-5

Operational Suitability	Tests
Sortie Generation	S-13
Sustainment	S-14

Table 2. USAF OT-IID COIs

Each COI is divided into multiple distinct Measures of Effectiveness (MOE) and Measures of Suitability (MOS), as appropriate. An MOE or MOS is a subset of the COI

that is “designed to treat specific and addressable parts” of the COI [Ref. 6:p. 13-7].

Each MOE or MOS is in turn divided into Measures of Performance (MOP), which are specific, well-defined measures that can be directly evaluated by a pilot or other test participant. At the bottom level of the hierarchy, each MOP may in fact have several specific questions that address it. Table 3 is an example of the hierarchical structure of the evaluation process, in this case the Assault Support COI, designated E-1 [Ref. 7].

The means for assessing specific MOPs are of two general types: objective and subjective. An objective measure is defined as one that is “not affected by personal feelings or prejudice; based on facts; unbiased.” [Ref. 8] A subjective measure, on the other hand, is one that is based on one’s opinion or attitude. As an example of an objective measure, Mission Capable rate in the Availability COI (S-3) is strictly a function of airframe uptime and downtime [Ref. 7:p. 5-6]. There is very little room for interpretation of the raw maintenance data that is used to determine these numbers. But by far the majority of MOPs are subjective in nature. For example, to assess the Fast-Rope MOP subordinate to the Aerial Delivery MOE in the Assault Support COI (E-1), the test participant must gauge numerous aspects of the effectiveness of the aircraft in performing the fast-rope mission, all based upon his experience and professional judgement. Although several specific questions may combine to address this rather broad MOP, the subject matter does not lend itself to precise, objective measurements, so the opinions and observations of the test participants must be relied upon to evaluate aircraft effectiveness in this mission area. While objective data collection certainly presents challenges, this thesis will focus on the collection and utilization of subjective test data.

COI/E-Test	MOE	MOP
Assault Support		
	Combat Assault Operations	Amphibious Pre-Assault/Raid Operations Amphibious Troop Lift Land Assault Troop Lift MOUT Maritime Intercept Operations GOPLAT
	Aerial Delivery	Fast Rope Rappel SPIE Helocast Parachute Ops/Personnel Parachute Ops/Equipment
	Air Evacuation	NEO CASEVAC
	TRAP	
	Air Logistical Support	External Payload Amphibious Assault External Cargo Land Assault External Cargo Internal Cargo
	Aircraft Refueling	Aerial Refueler Rapid Ground Refueler
	Battlefield Illumination	
	Assault Support Common	Mission Radius Cruise Airspeed Mission Planning Navigation Digital Map VSTOL Rough Terrain VTOL Litters Hoist Operations Defensive Weapon NVG HUD/HMD IR Sensor System One Engine Inoperative Operations Scramble Launch Short Notice Launch Wire Strike Protection Aerial Refueling Rapid Ground Refueling FARP Ground Refuel Defuel Fuel Dump

Table 3. Assault Support COI/MOE/MOP Hierarchy

The manner in which questions are generated will be discussed in the next section. But because in most cases each MOP is assessed through multiple questions, there are a large number of questions that must be answered in order to evaluate the six effectiveness and fourteen suitability COIs. In addition, each question must be answered

a sufficient number of times by a wide enough cross-section of participants and in a variety of environments and scenarios to allow conclusions to be drawn with a sufficient level of confidence. This creates a somewhat complex task with regards to data collection, data management, analysis, and reporting. The V-22 Questionnaire Tool (VQT) was created as an attempt at automating much of this task.

B. THE ROLE OF VQT

The manner in which the MOTT analysis section has elected to collect the vast majority of subjective test data is through the administration of post-event questionnaires. Each questionnaire is tailored for a specific test event and for the specific role the evaluator played in that test event. For example, an event that involves an external lift with a Cobra escort would require unique questionnaires to be generated for several groups of evaluators: the V-22 pilots, the V-22 enlisted aircrew, the Cobra pilots, and the ground support personnel handling the external load. (V-22 maintenance personnel performing maintenance on the aircraft prior to and following the event are also administered a questionnaire, but it is not event-specific).

Given that a number of specific tasks can be accomplished in a single mission/flight, that each task may be addressed by multiple questions, and that in a fashion typical for an operational environment the mission plan is highly subject to change at any time, the questionnaire construction process is highly dynamic. For each questionnaire, the questions appropriate for the task(s) and evaluator must be determined, collected, and sequenced in a logical manner. With prior knowledge of the testing schedule, this can be planned for and accomplished well in advance of the test event.

However, in order to maximize the efficiency of data collection, the process by necessity must be flexible enough to incorporate last-minute changes to the mission plan. Then once the questionnaires have been administered and data collected, the data must be aggregated, summarized, analyzed, and reported. VQT successfully automates much of this process.

The program was created by TSgt. Caldwell Roberson, USAF, the senior enlisted member of the MOTT analyst section. VQT is a Microsoft Access 97 database application that is intended to automate the questionnaire creation, administration, and summary reporting processes. TSgt. Roberson based VQT on a much simpler program provided by the Air Force Operational Test and Evaluation Center (AFOTEC) called Automated Rating Tool (ART), which was created for the purpose of generating questionnaires for the Air Force [Ref. 9]. Because ART did not quite fulfill the needs of the MOTT, TSgt. Roberson used his extensive knowledge of Access and Visual Basic for Applications (VBA) to create an application that was much more versatile and powerful. His use of a tailored relational database such as Access for data collection enables vast quantities of information to be stored and retrieved in an efficient manner. And indeed, in the two months covering the OT-IID test period, well over 23,000 responses to individual questions were collected by VQT-generated questionnaires.

The concept behind VQT is that in order to create a questionnaire, the user can input several items, such as mission area and participant category, to be used as parameters in a query, which will quickly and easily retrieve questions from the database that are relevant to the mission at hand. He can then present them on a preformatted

electronic questionnaire to the event participants. Once the questionnaire is administered and the data is entered into the database, it can be quickly and easily retrieved or manipulated in whatever fashion is necessary for analysis. A preformatted report is available that summarizes the data by COI, MOE/MOS, and MOP, for use by the Operational Test Directors (OTD).

1. Questionnaire Creation

Within VQT, the most basic element of data upon which the program functions is the question. There are several hundred unique questions in the question table, each one of which applies to one or more categories of participant and may apply to multiple MOPs. The questions themselves are created through a somewhat ad-hoc process, a series of consultations between the MOTT OTDs and the analysts. Each OTD is assigned one or more COIs for which he is responsible over the duration of operational testing. For each COI, the OTD and analyst work together to create questions that are designed, through a combination of operational experience and analytical reasoning, to concisely and completely address each MOP subordinate to that COI. It is a somewhat dynamic process, since unforeseen areas of concern arise during the course of operations and questions are created “midstream” as necessary to document the appropriate information. The MOTT is currently in the process of reviewing all questions utilized in OT-IID for revision and optimization as necessary prior to the initiation of OPEVAL [Ref. 9].

Each question is related within the database to one or more MOPs. Each question is also related to a single “Mission Area”, which is an aspect of the operational

environment, an aircraft hardware item, or a specific mission the V-22 is required to perform. One "Mission Area" may be related to numerous questions. The question/MOP relationship is established to facilitate database queries for reporting purposes. On the other hand, the "Mission Area" is not directly related to the COI hierarchy but provides a means by which a user can create a questionnaire by querying the database in terms of the operational specifics of a planned mission.

Figure 1. Questionnaire Editing Form

In order to create a new questionnaire, the user selects "Add a new Questionnaire" on the form presented in Figure 1. The user provides the necessary information, names the questionnaire, then chooses "Make questionnaire from library." Figure 2 illustrates the next screen the user is presented, by which he designates "Mission Areas" used to create a questionnaire. In this example, the user has selected "AIE/Aerial Delivery" and

“Air Logistical Support” as two of the mission areas that will be included on the questionnaire.

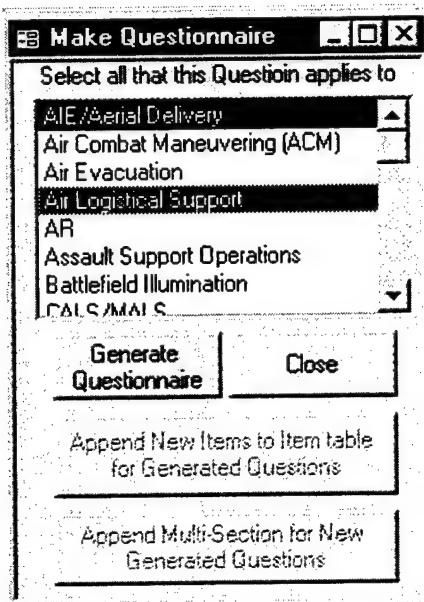


Figure 2. Mission Area Selection Form

Upon selection of the “Generate Questionnaire” button, VQT then queries the question database for all questions associated with chosen mission areas, and creates a new questionnaire based on the results. The new questionnaire is not immediately presented to the user; rather, it can be accessed and modified as necessary through an edit function.

Questionnaires created by the program, unless heavily edited by the analyst prior to administration, have demonstrated a tendency to be excessively long. For example, several of the questionnaires created for administration to MOTT pilots during OT-IID are in excess of 80 questions. The reason for this lies partly in the manner in which the questions are retrieved from the database for inclusion on a questionnaire. The query by which the program retrieves all mission-relevant questions desired by the user uses

“Mission Areas” as the sole criteria for determining what records to include in the questionnaire. Therefore every relevant record is retrieved without regards to the length of the resultant questionnaire or the quantity and characteristics of previously collected data. This becomes a problem in the case of single missions that cover multiple mission areas, in which case the number of relevant questions in the database increases rapidly. An individual participating in the mission is then required to complete an excessively long questionnaire at the conclusion of that test event. In this type of instance, the quality, accuracy, and completeness of the responses that are elicited may be questionable. In addition, without examining previously collected data, a very uneven manner of data collection may arise as the more common mission areas are repeatedly covered by questions, while the more obscure areas are queried infrequently or not at all.

The questionnaire administered to the maintenance personnel is a unique case. Because of the unplanned nature of organizational maintenance, a standard maintenance questionnaire was created for administration following maintenance actions. The questionnaire used during OT-IID consists of 17 questions and is generic enough that it can be administered for all maintenance actions. The unavoidable byproduct of this approach is that in virtually every questionnaire administered to a maintenance Marine there are a number of questions that are not applicable to that maintenance action.

2. Questionnaire Administration

Once the questionnaire has been created, it is available to the event participants through a separate interface designed for administering questionnaires. Figure 3 is an example of the form provided through which an event participant completes a

questionnaire. The individual logs into the network and starts the program, chooses the appropriate event from the list provided, specifies his name and some additional

30. Rate the effectiveness of the mission planning systems interaction with V-22 DIGMAP.

MP03

Extremely Ineffective Moderately Ineffective Barely Ineffective Barely Effective Moderately Effective Extremely Effective N/A or Not answered

Areas to consider (Please select all that Apply)

Area coverage
Clarity
Map symbology presentation

Comments

Questionnaire Control #: 000001976 Question: 30 of 56 Question Answered

Prev Quest Next Quest Close

Figure 3. Questionnaire Administration Form

information, and proceeds to complete the questionnaire. The maintenance department is the notable exception. Due to the lack of network terminal availability for the large number of maintenance personnel involved, a procedure was implemented by which maintenance personnel are administered a paper copy of their questionnaire, which is collected at a central location and periodically input into VQT by a maintenance representative.

By using user-friendly forms and controls, restrictive program access, and by making VQT resident on a local area network, data collection is greatly facilitated since event participants can simply log into the network and fill out the questionnaire that has

been created for them. Thus the data is entered directly into the database by the evaluator, with no intervention necessary by an intermediary. All data is stored with information specifying the evaluator, the event, and other information to facilitate retrieval for analysis. The option to generate a paper copy of the questionnaire is available for administration to non-local participants.

3. Summary Reporting

A preformatted report option has been built into VQT to enable an analyst to create summary reports on the data collected to that point in time. A report is generated for each COI specified by the user. Figure 4 is an excerpt of an example report that may be produced. While the contents are very straightforward, the format of the report bears some explanation.

The report summarizes the data at the question level, and the questions are grouped by COI, MOE/MOS, and MOP. There is no aggregation of data above the question level. Many of the questions are presented at the COI and MOE/MOS level, rather than at the MOP level. The explanation for this is as follows: because many of the questions apply to multiple MOPs, those questions were relegated for reporting to the MOE/MOS or COI level, as appropriate, to minimize redundancy in the generated report [Ref. 9]. For example, if a question concerning the AN/AVS-7 HUD was determined to apply to every MOP subordinate to the Combat Assault Operations MOE, that question would be summarized once in the report under that MOE, rather than multiple times, once under each of the applicable MOPs.

E-3 Results

E-3-3-3

HF10: Rate the effectiveness of the environmental control unit to maintain a suitable crew and cabin occupant comfort level throughout the operating environment.

	Extremely Ineffective N/A 33	Moderately Ineffective (1) 108	Barely Ineffective (2) 21	Barely Effective (3) 16	Moderately Effective (4) 45	Extremely Effective (5) 39	Total Responses (6) 22
--	---	---	--	--------------------------------------	--	---	-------------------------------------

Comments

Event 1 (Air Crew)

Cockpit is good when the ECS works, cabin is unacceptable, to hot!!!!

Response Respondent

1 TSgt Welk

This is a very serious problem due to the inadequate ability of the ECU to circulate and send cool air to cabin. The lack of circulation is probably the most serious problem. In APLN mode with all doors closed this makes for a very miserable flight also physically draining.

1 Sgt Banks

Nearly totally in-effective. Does little to effectively cool the cabin, especially on warm or hot days. It provides no fresh air.

4 SSgt Huntington

The only thing it does provide is warm to ambient recirculated stagnant air to the cabin air.

TOTALLY UNACCEPTABLE

1 Sgt Hurley

Event 1 (Pilot)

Works ok in the cockpit only. Even with a fully functional ECS, cabin ambient temperatures are extremely high even in cool weather. **WILL NOT BE ACCEPTABLE IN DESERT/JUNGLE ENVIRONMENT.**

3 Maj Westman

cockpit temperature remains comfortable, but the ecs is incapable of keeping the cabin comfortable with 85 degree ambient air

5 Lt Col Shaffer

We need a better system for our troops in the cabin, the aircrewmen will begin to fatigue much sooner than the pilots, and the fighting force in the cabin will be ineffective, once in the objective area, because of the hostile environment in the cabin.

1 Maj Torres

The ECU functioned well in the cockpit for the entire flight. It was still hot in the cabin area.

5 Lt Col Sweaney

Event 10 (AIE troops)

nce

N/A LCpl Franco

was a little warm.

5 LCpl Gray

Must be better than if the airplane didn't have it at all. it was still hot but bearable. I guess I'm not use to anyone calling Marine air a comfortable thing.

N/A SSgt Colonna

I can't see this bird being anywhere near as versatile as a ch-46.. As a corpsman, I don't see being able to transport patients as well as marines simultaneously as we often have to.

4 HA (Navy) Murray

Hot inside a/c with aft ramp & hatch secured; however with the ramp door open & upper crew door open temp was comfortable.

4 Sgt Starzynski

No idea

N/A GySgt Blackburn

To hot. Need to moderate heat.

5 LCpl Clynes

Page 28 of 46

Figure 4. Example Page of VQT Summary Report

There is no automated data analysis performed on the response scores beyond the textual histograms presented. The primary consumers of the reports generated by VQT

are the OTDs, who use the information presented, supplemented by their professional judgment and experience, to evaluate the V-22 in terms of the COI(s) they have been assigned. These individuals are not trained in statistics, yet must incorporate with their own insights the vast amounts of data collected and generate an accurate and coherent evaluation report for submission to Commander, Operational Test and Evaluation Force (COMOPTEVFOR). The reports that VQT generates to summarize the data, although comprehensive and easy to interpret, are fairly rudimentary and in most cases excessive in length. For example, the summary report for the Assault Support COI (E-1) following OT-IID was 505 pages.

While the current report contains all the pertinent information for that COI collected over the test period, it is difficult, if not impossible, to build an awareness of the overall test period results solely through the use of this report. As can be seen in Figure 4, all information is presented textually, and the sheer size of a 505-page summary and quantity of information presented is overwhelming. Without tremendous effort on the part of the OTD, this would not seem to readily facilitate the overall insight of the OTD with regards to the COI and subordinate measures he must address. The OTD should certainly be familiar in detail with the data being collected, as it involves not only numbers, but also observations and insights, and all details must be made available. But a means to summarize the data in pictorial or graphical form on a macro level was not written into VQT.

III. PRELIMINARY ISSUES

A primary concern of this thesis is that any product developed be practical, easily interpreted, and help ease the burden on the OTDs. There is also the implicit expectation that any method used or assumption made be sound and grounded in established statistical theory. The ability to satisfy both demands is not always a simple task, and the issues involved are not necessarily clearly defined.

This chapter discusses several preliminary issues that must be resolved prior to implementing any additional functionality within VQT. The first of these issues is the computing environment in which any software solution is to be implemented, and the programming technique to be utilized. Next, the original construction of the database is examined in terms of efficiency and size, and the implications those factors may have for OPEVAL data collection. And third, the issue of measurement and classification of data types is discussed, with the objective of identifying an appropriate measure by which to summarize and analyze numerical questionnaire data.

A. PROGRAMMING ENVIRONMENT

1. Software

Microsoft Office is becoming the standard office productivity suite in use throughout the Navy and Marine Corps. Because VQT is an application of Access, the Microsoft Office database, any programmatic enhancements or additions to VQT should remain within the confines of the Microsoft Office suite. In this way, the program remains readily transportable and easily installed, and the interface is relatively familiar to any potential Navy or Marine Corps user. This necessarily precludes the use of more

powerful and capable programs, such as S-Plus or Minitab, for conducting any sort of automated analysis within VQT.

2. OLE Automation

A distinct advantage of working within the confines of the Microsoft Office suite is that object linking and embedding (OLE) can be put to use with relative ease. OLE offers the user, through the user interface, the “ability to expose the object model used in other applications and then manipulate those objects in much the same way you manipulate the objects in the client application.” [Ref. 10:p. 382] OLE permits the user to establish links from embedded objects back to the object’s source application, thereby permitting the use of that source application to create and edit the object.

Similarly, and perhaps more importantly in this context, the power of OLE Automation can readily be used to expand the utility and capabilities of the basic VQT application, through the use of programming tools which are inherent to the applications of the Microsoft Office suite. OLE Automation, or more recently known as just Automation, “allows you, the developer, to link to objects (such as documents, spreadsheets, and so on) from other applications, or embed objects from other applications in forms, reports, or tables, thereby allowing the native application to perform the tasks needed.” [Ref. 11:p. 346] In short, Automation allows the Visual Basic for Applications (VBA) programmer to remotely control the actions of a program, for instance Excel, through code within the context of another program, such as Access. This is accomplished through the use of application object models and object libraries,

which are exposed to the VBA programmer through the Visual Basic Editor and enable him to manipulate the methods and properties of objects belonging to other applications.

The value of such a tool becomes apparent when one examines the strengths and weaknesses of the individual applications within the Office suite. For example, Access is a powerful relational database that lends itself to the storage and manipulation of vast quantities of data. It is, however, limited in its ability to do statistical analysis or present data graphically. On the other hand, Excel is a spreadsheet program that is a very versatile and powerful tool for conducting numerical analysis and incorporates a highly diverse capability to present data graphically. Excel is limited, however, in the quantity and complexity of data it can store and readily manipulate. The capabilities of these two programs complement each other, and Automation provides a tool for accomplishing a relatively seamless and user-friendly integration of the two applications.

B. DATABASE NORMALIZATION

One advantage to using a relational database, such as Access, lies in its ability to store vast quantities of complex data, while not incurring the penalty of redundancy inherent in maintaining such records in a flat database, such as a spreadsheet [Ref. 12:p. 3]. Because the data collected in the operational test environment by nature contains massive redundancies, a relational database is a natural choice for data management. Examples of these redundancies include names and ranks of test participants, questions presented, test event specifics, etc... All information must be, and is archived in such a way as to maintain original relationships and allow re-creation and appropriate manipulation of the data for analytical purposes. If properly designed, a relational

database facilitates accurate data storage while minimizing redundant entries. The process of changing a database structure to comply with this design goal is called normalization [Ref. 12:p. 38].

A close scrutiny of the structure, relationships, and functional code within VQT reveals that the database is in need of normalization prior to implementing any further enhancements to the program. Although VQT is functional as it stands, several relationships and subroutines exist which create additional redundancies each time they are called. For example, VQT contains a "Questions" table that is populated with all the questions used to create questionnaires, and 13 other informational fields (items) associated with each question. A question together with its associated fields is known as a record. Although the table contains only 406 unique questions, the actual size of the table, by the conclusion of OT-IID, the first test period in which VQT was fully utilized, had blossomed to 4593 records.

This growth is the result of an event handler that creates copies of full records from the "Questions" table as a means to recreate questionnaires. These duplicate records are appended to the end of the "Questions" table and assigned the appropriate questionnaire number as a link between questions and questionnaire. Although this design problem led to no particular difficulties during OT-IID, it was a contributor to the gradual growth of the database to an unnecessarily large size by the end of the test period. OT-IID was a two-month operational assessment (OT-IID) involving two aircraft and six MOTT pilots. By comparison, OPEVAL is a much larger evolution scheduled for approximately seven months, and will involve four aircraft and thirteen MOTT pilots.

Because there are several instances of inefficiencies similar to the one described, the potential exists for problems to arise during an extended test period such as OPEVAL, in the form of an unnecessarily large, unresponsive, unwieldy database that would worsen as the test proceeds.

Therefore, prior to adding any additional functionality to VQT, an effort is made to normalize the database, at least to the extent that future growth of the database will be primarily attributable to the collection of new data. This effort involves the addition of several tables, a small amount of restructuring of database relationships, and the re-coding of several event handlers to account for the new structure. For the reader interested in database structure, the relationship diagrams for the old and revised versions of VQT are provided in Appendix A.

The altered version of VQT uses, as a basis, the program as it stood at the completion of OT-IID in October 1998. Because it is populated with actual test data from OT-IID, all changes are made under the imperative of preserving that data as well as all previous functions of the program. The changes in structure and subsequent cleanup of unnecessary redundancies actually reduces the size of the database from 7.3 megabytes, its size in November 1998, to just over 5.3 megabytes. As a reference point, actual test data collected comprises approximately 2.8 megabytes of the database, while the basic program is approximately 2 megabytes. Thus the "overhead" involved in using VQT, which consists of peripheral information that is generated as the collection of usable data grows, is decreased from roughly 2.5 megabytes to roughly 0.5 megabytes. Full functionality and the integrity of the data are maintained. These figures do not

reflect the fact that the size of the revised structure will grow at a much slower rate as compared to the old structure. This revised structure provides a stable and more efficient program on which to create further functions and capabilities.

Assuming that in both the original and revised programs the amount of “overhead” is a linear function of the amount of data collected, a crude prediction of the relative performance of the VQT versions in OPEVAL can be made. As Table 4 shows, the amount of usable data collected during OT-IID was approximately 2.8 megabytes. When the starting size of the program and the amount of usable data are subtracted from the size of the program with data, what remains is the “overhead.” Comparison of the

Version	Size with data	Usable OT-IID data	Base program size	Overhead generated	OH/Data Ratio	OPEVAL prediction
Original	7.3M	2.8M	2.1M	2.4M	0.86	28.1M
Revised	5.3M	2.8M	2.0M	0.5M	0.18	18.5M

(All quantities are in megabytes)

Table 4. Comparison of Relative Efficiency Between VQT Versions

ratios of “overhead” to usable data for the two versions reveals that, by this measure, the revised version is nearly five times as efficient as the original. OPEVAL is scheduled for approximately five times the flight time that was flown in OT-IID [Ref. 4:p. IV-8]. If flight time is assumed to be a linear predictor of the quantity of data (and therefore “overhead”) that will be collected, the predicted database sizes can be calculated as a simple multiple of the OT-IID numbers. As the last column of Table 4 shows, the revised version of VQT should result in a substantially smaller database.

C. MEASUREMENTS

In attempting to provide a meaningful, accurate, and useful summary of the accumulated numerical ratings, the related issues of data types and statistical measures

must be addressed. The type of data being examined determines what types of statistical summaries are appropriate, in addition to determining what types of subsequent statistical analysis are appropriate or even possible. This becomes a central topic whether discussing the screening criteria for questionnaire building, alternate methods of data presentation, or appropriate analytical tools to use on the data set.

The V-22 Test Plan [Ref. 7:p. D-2] discusses the manner in which numerical test data is to be handled. The "quantitative subjective" data to be gathered from the questionnaires is administered on what is considered to be an equal-appearing intervals scale. The respondent chooses from a list of six alternatives, each being assigned a value from one (worst) to six (best). Refer to Table 5 below, for the responses and their

Response	Definition	Value
Extremely Ineffective/Unacceptable	Task cannot be performed or the item is unusable or unsafe. Mission/task not accomplished due to equipment deficiencies or procedural limitations	1
Moderately Ineffective/Unacceptable	Major problems encountered. Task accomplished with great difficulty or accomplished poorly. Significant degradation of mission/task accomplishment or accuracy	2
Barely Ineffective/Unacceptable	Minor problems encountered. Task accomplished with some difficulty. Some degradation of mission/task accomplishment or accuracy	3
Barely Effective/Acceptable	The item or task meets its intended purpose with some reservations. Meets minimum requirements to accomplish mission or task	4
Moderately Effective/Acceptable	The item or task meets its intended purpose; it could be improved to make it easier or more efficient	5
Extremely Effective/Acceptable	The item or task is clearly satisfactory the way it is; no improvement required	6

Table 5. Questionnaire Response Definitions

definitions. Therefore, according to the Test Plan, this data is suitable for treatment as interval data. The text goes on to state that the

...distance between each value is equal and predictable. Therefore, the respondent population will understand (at least subconsciously) that each point of the adequacy scale is equally spaced, like the known and equal intervals on, for example, a thermometer...As a result of the above

rationale, the six-point scales can be manipulated and analyzed to produce statistical means and standard deviations. [Ref. 7:p. D-3]

There is a large body of literature available on the topic of measurement scales and data types; see Agresti [Ref. 13:pp. 2-4], for example, for a concise discussion on the topic. Whether the data should be considered ordinal or interval in this case depends heavily on the degree of standardization among the test participants, a consideration that the response definitions given in the Test Plan attempt to address. The viewpoint taken will necessarily determine what statistical tools are available and appropriate for examining and presenting the data.

It is impractical and of limited use to display frequency distributions for the responses to each question. Thus, at issue is the basic question of which summary statistics have the ability to adequately summarize the data for the task at hand. An OTD or analyst must be able to determine whether the magnitude of the responses indicate that the item in question is deemed acceptable, and whether the spread of the responses indicates some level of consensus among respondents. Given a typical distribution, such as the Normal distribution, the mean and standard deviation provide sufficient information to facilitate a decision. In this case however, the distribution of responses is discrete rather than continuous; it is confined to the integer values one through six; it is in some cases bimodal; and it is in almost all cases highly skewed. Therefore, two summary statistics describing central tendency and spread cannot sufficiently capture all the important features of the response distributions.

Figure 5 provides examples of the types of distributions that may be encountered with this type of data, and the associated means, medians, standard deviations, and interquartile ranges. Each histogram plots 100 values, and all statistics were generated with S-Plus 4.5. Plot six is probably unlikely to occur with the VQT questions, but is presented for comparison purposes.

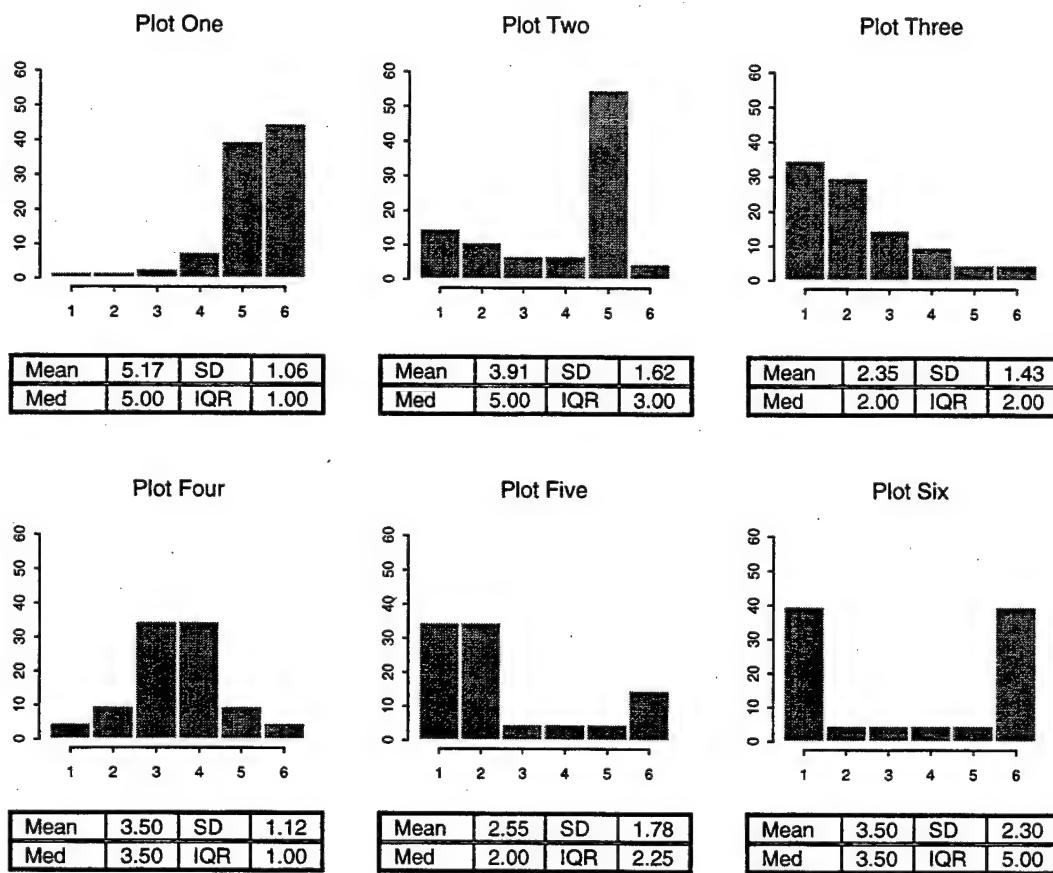


Figure 5. Histograms of Example Questionnaire Data Distributions with Corresponding Means, Medians, Standard Deviations (SD), and Interquartile Ranges (IQR)

In plots one and four, there appears to be a fairly strong consensus among responses, and the two measures of central tendency and spread are in agreeance. Notable differences occur in the bimodal situations, plots two, five, and six. The heavy

left tail of plot two pulls the mean substantially away from the median. The standard deviation and the interquartile range of this plot are very different, although both reflect a high degree of spread. That difference is accentuated in plot six, an extreme case. Plots three and five are the least extreme cases, and result in differences between measures that are noticeable, but not extreme. In general, from this example it is clear that measures of spread and central tendency do not always agree and that two measures are probably not adequate to summarize all possible data of this type.

To investigate this issue in the context of actual test data, a comparison is made of the information provided by three different methods of measuring the “adequacy” of responses to a question. The first uses the mean and standard deviation, another uses the median and interquartile range, and the third uses a measure described in the Test Plan as the “80/50 Rule”. The 80/50 Rule states that for the results of a particular question to be deemed “acceptable”, a minimum of 80 percent of the responses to that question must be at least a four, and at least 50 percent must be a rating of five or six [Ref. 7:p. D-6]. In more formal terms, the 20th percentile must be at least a four, and the median must be at least a five. Because it is the method specifically described in the Test Plan, the 80/50 Rule is used as a baseline for the other two measures.

In order to conduct the comparison, OT-IID test data is extracted from VQT in the form of all numerical responses to the first 200 of the 406 distinct questions in the database ordered alphabetically, and placed on an Excel spreadsheet. This number is chosen to remain within the column limits of an Excel spreadsheet. For each of the 200 questions, the responses are summarized using each of the three methods in question, and

pass or fail assigned by each method. The threshold values for mean, standard deviation, median, and interquartile range used as pass/fail criteria are then adjusted to produce roughly the same number of failures as the baseline 80/50 Rule. In addition, only questions with a minimum of 15 responses are passed. All computations are accomplished in Excel 97.

		(Mean \geq 4, SD \leq 1.4)	
		Pass	Fail
(80/50 Rule)	Pass	131	4
	Fail	5	60

Table 6. 80/50 Versus Mean & Standard Deviation

		(Median \geq 4, IQR \leq 1.5)	
		Pass	Fail
(80/50 Rule)	Pass	132	3
	Fail	6	59

Table 7. 80/50 Versus Median & Interquartile Range

		(Median \geq 4, IQR \leq 1.5)	
		Pass	Fail
(Mean \geq 4, SD \leq 1.4)	Pass	131	5
	Fail	7	57

Table 8. Mean & Standard Deviation Versus Median & Interquartile Range

As can be seen in Tables 6, 7, and 8 by the large numbers on the diagonal of all three tables, the three methods being compared concur for the vast majority of questions summarized. This comparison indicates that the 80/50 Rule, the mean and standard deviation, and the median and interquartile range provide very nearly the same information in terms of what is required to evaluate this manner of test data. It should be

noted that because the responses are subjective, they will in every case be augmented by the textual remarks provided by the evaluators and the personal experiences and professional opinion of the OTD.

In light of the results of Tables 6-8, the mean and standard deviation are chosen for use as the measures of central tendency and spread. There are two advantages to using these measures. First, the MOTT analyst section maintains that the mean and standard deviation are fairly intuitive measures that can be more readily interpreted by the OTDs than the median and interquartile range [Ref. 9]. And second, because Microsoft Access and Microsoft Excel are the working environment chosen for VQT and its enhancements, use of the mean and standard deviation is much more convenient for programming, data manipulation, and summary presentation. These programs do not provide the flexibility and analytical power of other statistical packages such as S-Plus. Access does not provide the ability to calculate the median or quartiles on a data set, and although Excel does, it is not able to handle every data set to be summarized due to spreadsheet dimension limitations. So the use of the median and quartiles would require substantially more data manipulation and programming, and would not result in any clear benefit.

IV. PROGRAMMING METHODOLOGY

The effort to program additional functionality into VQT has two objectives; to provide an automated process for generating shorter questionnaires, and to make available a more effective data summary presentation. Both are primarily exercises in VBA programming, and make use of some basic summary statistics. This chapter is a description of the general processes and methodologies on which the VQT design changes are based.

A. CRITERION-BASED SCREENING

The issue of excessive questionnaire length is one that is of primary concern. The desire for sufficient sample size and thorough coverage of each COI would seem to be at odds with an attempt to limit the length of the questionnaires. Because of the dynamic and diverse nature of operational testing, the complexity of some missions requires that the corresponding questionnaires be substantially longer than others. Therefore setting an arbitrary limit on the number of questions assigned to a questionnaire is not appropriate.

The approach chosen for this thesis is to provide the user with an option to pre-screen questions based on some criteria. These criteria screen each question in terms of the responses already collected, prior to including each question in the questionnaire he is creating. The idea is to continue to gather data on one particular question only as long as necessary to reach some degree of confidence in the answer provided by the data. That degree of confidence is dependent on criteria provided by the user. By using these criteria as parameters in an appropriate query, VQT can dynamically and automatically

determine if each question pertinent to the test event being conducted should continue to be presented, or if it can be justifiably omitted from consideration and thereby shorten the questionnaire.

The criteria chosen for automated screening, for reasons discussed in Chapter III, are the sample size, mean, and standard deviation of the numerical responses to the question. In order to make comparisons with the user-provided criteria, VQT first must compute the summary statistics. This is facilitated by the use of a “Make-table” query that is run every time the user initiates the screening process. Specifically, it is initiated by the event handler that is fired when the user clicks on the “Create questionnaire” button on the “Screening Criteria” form. A “Make-Table” query is, as the name implies, a query by which Access creates a “base” table, fills it with the results of the query, and stores it. This is contrast to a “derived” table, the results of a typical query, which are not stored [Ref. 12:p. 52]. For each of the questions contained in the QUESTIONS table, VQT queries the ANSWERS table and computes the mean, standard deviation, and number of responses recorded for that question. Responses of “NA” are not included. A new table is then created (overwriting any previous version) containing the results of the query. The summary statistics of each question are then available for comparison with the threshold values that the user has provided. Questions that meet user criteria are excluded from the questionnaire, and all others are included. The following paragraphs walk the reader through the user interface.

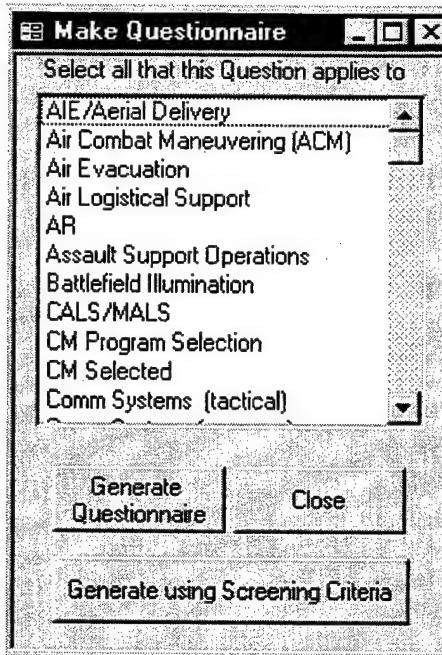


Figure 6. Revised Mission Area Selection Form

A user initiates the creation of a questionnaire in the same manner as described in Chapter II. Upon reaching the form on which he designates his pertinent mission areas though, he is now presented with an additional, new button entitled "Generate using Screening Criteria". Figure 6 illustrates the new form. Once he has designated his mission areas, he has his choice of options to pursue. He can choose to bypass the screening process by clicking on the old button, "Generate Questionnaire", and initiate the basic questionnaire generation process as described in Chapter II. If, however, he chooses to designate criteria by which VQT will additionally screen the questions within the chosen mission areas, he clicks on the new button.

Figure 7 is an example of the form the user will then be presented. Here the user is given the opportunity to provide criteria for threshold mean, standard deviation, and sample size values. The idea is that the user wants to exclude from the questionnaire any

questions whose previous numerical responses have an average value greater than or equal to, a standard deviation less than or equal to, and a sample size of at least as great

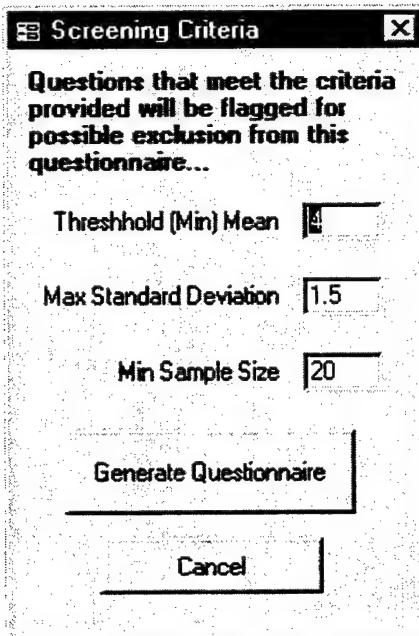


Figure 7. Screening Criteria Selection Form

as the threshold value given. The values shown are default values, which have been set to render roughly the same results as the 80/50 Rule, discussed in Chapter III. These values are intended to be overwritten, but if the user chooses not to do so, VQT will conduct the screening based on the default values.

Once the user clicks on “Generate Questionnaire”, the event handler is fired that creates the table of response statistics and, for each mission area that was previously designated by the user, compares statistics on the all pertinent questions to the criteria provided. Each question that meets all criteria is excluded from the questionnaire and referenced in a table of excluded questions. All others are included on the new questionnaire.

The results of the comparison are presented to the user, in a manner similar to Figure 8. This form shows the user exactly which questions were included in the questionnaire and which were excluded, and provides the option for the user to override the query and place excluded questions back onto the questionnaire. Because references

Screening Results : Form

Included Questions (33 total)	
AS01	Rate the effectiveness of conducting AMPHIBIOUS PRE-ASSAULT/RAID operations.
AS02	Rate the effectiveness of conducting AMPHIBIOUS TROOP LIFT operations.
AS04	Rate the effectiveness of conducting MOUT operations.
AS05	Rate the effectiveness of conducting MIO operations.
AS06	Rate the effectiveness of conducting GOPLAT operations.
CAL01	Rate the effectiveness of landings in a confined/mountainous area of varying topography.
CAL03	Rate the effectiveness of aircrew clearing the aircraft of all obstacles during landing and takeoff

Criteria: Mean ≥ 4
Standard Deviation ≤ 1.5
At least 20 responses

Excluded Questions (7 total)	
AS03	Rate the effectiveness of conducting LAND ASSAULT TROOP LIFT operations.
CAL02	Rate the effectiveness of forward and downward cockpit visibility during CALS/MALS.
COMT02	Rate the suitability of the Troop Commander's station to access the aircraft's radio.
FS01	Rate the overall safety during aircraft takeoff.
FS02	Rate the overall safety during enroute operations.
FS03	Rate the overall safety during terminal/objective area operations.
FS04	Rate the overall safety during landing.

Accept Questionnaire **Re-add selected Question(s) to Questionnaire**

Figure 8. Example of Screening Results Form

to the excluded questions are maintained in a table, it is a simple matter to recall them or replace them on the questionnaire. If the user is not satisfied with the results of the screening, he may highlight one or more excluded questions, click on "Re-add selected Questions to Questionnaire", and the selection(s) will be added to the questionnaire, appropriate references in the excluded questions table being removed. The contents of the form are immediately updated to reflect the change(s). This may be done repeatedly.

Once satisfied with the questionnaire, the user selects “Accept Questionnaire”, and the process is complete. The ability to remove questions from the questionnaire is not provided on this form. This capability is incorporated into a questionnaire editing form that remains intact in this version of VQT.

The VBA code that accomplishes the screening function is provided in Appendix B. It consists of several subroutines, the most notable being “cmdGenerate_Click()”, the code that actually conducts the comparison of question statistics to user-provided criteria. The purpose of using the seemingly unnecessary make-table query vice a typical query can be seen by examining the context in which the table is referenced. The code is looping over a somewhat involved Structured Query Language (SQL) statement that, by necessity, references the results of a query. By running the query that saves the results in a table prior to entry into the loop, rather than re-querying each time through the loop, the table can be referenced each time, thus saving substantial CPU time. So although initially the use of a make-table query seems to create unnecessary redundancy in the database, it does so with good cause.

B. REPORTING

The central objective in creating a summary report is to provide an easily interpreted yet highly informative presentation. Several of the COIs to be reported are aggregating thousands of individual responses in dozens of subsets, which creates a challenge in determining how the raw data is to be handled and how it is to be summarized. As discussed in Chapter III, the summary measures chosen for the data are the mean, standard deviation, and sample size. In order to be effective in its intended

role, a report of some or all of these summary measures must be presented in a manner in which the OTDs and analysts can quickly and effectively extract the information they require for monitoring test progress, planning, and formal reporting.

Numerous graphical techniques of varying degrees of sophistication exist for displaying this type of information. Within Microsoft Excel alone, there are dozens of predefined charts available and appropriate for this data as well as the capability for the user to define his own chart format. For example, a bubble chart and a chart of floating bars are each capable of displaying the same information, although in very different formats. The limiting factor in the case of VQT is the fact that the OTDs, the primary intended users of the summaries, are not necessarily well-versed in statistics and sophisticated graphical methods, yet need to pull salient features from the summaries. With this in mind, guidance from the MOTT analysis section provides a specific framework in which to develop a solution. The data to be presented to the OTDs is to be the mean score and number of numerical responses for each question, which together will provide sufficient information for their purposes. It is speculated that the standard deviation may result in too many dimensions on the chart, may not be well understood, and will therefore probably not be utilized [Ref. 9]. It should be noted however, that although the standard deviation is excluded from the summary charts, it is still readily available in textual format to the analysts. This will be discussed later in the chapter.

Because the decision was made early on to confine any VQT enhancements to programs within the Microsoft Office suite, Excel is the obvious choice for a graphical solution, due to the inherent flexibility, diverse range of pre-defined options, and ease of

use of its charting capability. A charting tool in the form of Microsoft Graph is provided within Access, but it does not present nearly the power or flexibility of Excel in that realm. The primary options considered for integrating the two programs are through OLE, by embedding summary charts created in Excel within an Access report generated by VQT, or with Automation, using VBA to create and present a summary report entirely within Excel. In order to preserve the original COI reports, which are still important to the OTDs as a comprehensive compilation of the questionnaire data, and in an effort to maintain simplicity as much as practical, Automation is chosen for the implemented solution. Through the use of Automation, as discussed in Chapter III, the integration of Access and Excel becomes a somewhat straightforward application of VBA programming, the result of which makes use of the strong suits of each of the programs.

In general terms, the graphic reporting function that is implemented takes data from Access, opens an application of Excel, inserts the data on a spreadsheet, creates a number of charts based on that data, and presents them to the user. The entire process is transparent to the user, and is initiated with the click of one new command button on a previously existing VQT form. The result is a series of graphs, presented in a format ready for printing, that summarize all data collected for a COI that has been selected by the user. The following paragraphs will present the process in more detail.

Figure 9 represents the form that a user is presented with when he clicks on the “Reports Menu” button on the “V22 Questionnaire Manager / Analyst Menu” form. Note that this is a limited access function that is available only to analysts and appropriately designated personnel. In the center section of the form the options are provided for

producing preformatted reports. The user who wants to create a graphical summary selects the COI of his choice from the listbox entitled "COIs", and clicks on the new

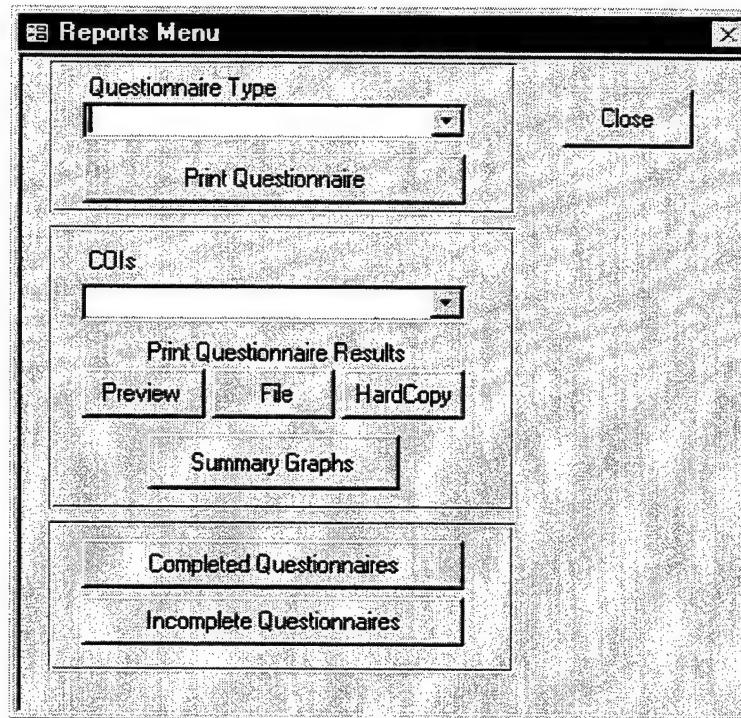


Figure 9. Revised Reports Menu Form

button labeled "Summary Graphs." At this point a series of events is initiated. The code for these subroutines, beginning with the event handler that is triggered by clicking the "Summary Graphs" button, is provided in Appendix C as a reference for the reader who desires more detail than provided here.

As the first event, VQT initiates a parameter query that references the COI chosen by the user, extracts all responses from the ANSWERS table that are relevant to the chosen COI, and calculates a number of statistics on those responses. These statistics include the number of numerical responses and the mean, standard deviation, and standard error of those responses. Each line in the resulting table consists of the statistics

on one question that is assigned to the selected COI or a subordinate MOE/MOS or MOP.

This is the data that is passed to Excel for charting.

Now begins the Automation, in which VBA subroutines remotely manipulate Excel procedures and methods from within Access. An application of Excel is opened, although it is not made visible to the user until nearly the end of the subroutine. It is provided a workbook with worksheets, and the data resulting from the previously accomplished Access query is pasted onto an Excel spreadsheet.

At this point, using a series of loops, the data is examined to determine how it is organized and how it will be broken up into separate charts. As an aside, because the size of the data set can vary widely from one use to the next, user-defined collection objects are a particularly convenient type of VBA object that are used here, as they are in essence the VBA version of a dynamic array [Ref. 11:p. 68]. The data is partitioned for charting by COI, MOE/MOS, and MOP, and further partitioned into groups of ten or less for placement on individual charts. The number ten is purely an arbitrary choice by the author as the maximum number of data points to place on a single chart, in an effort to minimize clutter and maximize readability of each chart.

The charts are then individually created, formatted, and placed on a second spreadsheet. A maximum of six charts are placed on a single spreadsheet, at which point a page break is forced prior to adding more charts. The Excel application becomes visible to the user just as the Print Preview function is called. So after the click of the “Summary Graphs” button, the next action the user views on the screen (after a slight delay) is the Excel print preview window, displaying his page(s) of charts summarizing

the data for the chosen COI. The subroutine ends when the user clicks either "Print" or "Close" in the print preview window, and the user is returned to the Reports Menu screen in VQT. Excel is left open so the user, presumably an analyst, has available the spreadsheet of raw data used for charting, in addition to the standard deviation and standard error for each question.

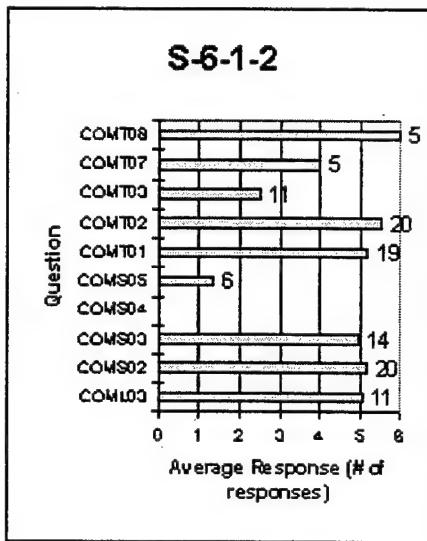


Figure 10. Typical Summary Chart

Figure 10 illustrates the layout of the charts that are produced. The title is the COI, MOE/MOS, or MOP to which these questions have been assigned, using a designation that is consistent with that contained in the Test Plan. Since one is the minimum value that an evaluator can assign to a question, a zero value portrayed on the chart signifies a question that has not yet been answered with a numerical response. The labels on the vertical axis are shorthand labels used by VQT to refer to each unique question, and must be referenced separately to determine the exact question wording. The numbers provided off the end of the bars are not, as might be expected, the bar value.

Instead they provide the user with the sample size, the number of numerical responses received for that question.

The entire series of these charts are produced and laid out six per page for printing. Figure 11 is an example of a page that may be produced. As can be quickly noted, it becomes a simple matter to pick out the questions that are not producing satisfactory response levels, either in terms of the mean response value or in the number of responses collected to-date. As the figure shows, even the questions that have yet to be answered are depicted, those being the cases with no bar present. Thus the OTD is provided with a complete summary of all questions pertinent to his COI.

One aspect of the code that may appear unusual bears some explanation. Due to the manner in which the OT-IID database questions are organized, some of the larger COIs require the creation of over 70 charts to summarize the data. While the number of charts required is dependent on how the questions are assigned to COIs, MOE/MOSSs, and MOPs, and this may change somewhat between test periods, it is to be expected that the OPEVAL summaries will be of similar magnitude to those for OT-IID. The code, it may be noted, after creating 24 charts on a spreadsheet in one Excel application, creates and opens another application of Excel to continue with the next 24 charts, rather than continue to produce more charts on the same spreadsheet. And the process continues until all necessary charts have been produced. The summary of the Assault Support COI (E-1) for the OT-IID data, for example, requires the creation of 75 charts, which results in four separate applications of Excel being created.

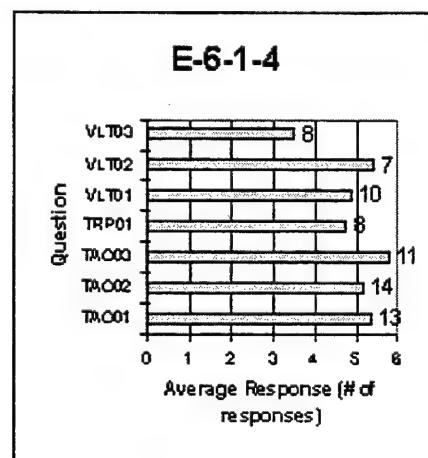
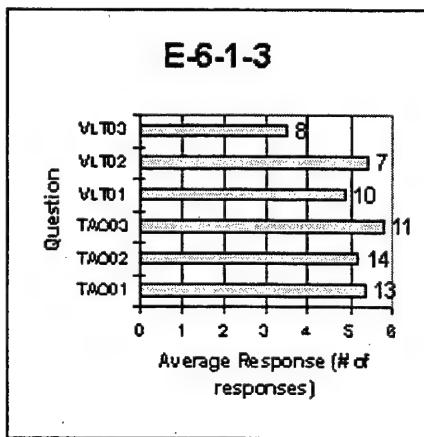
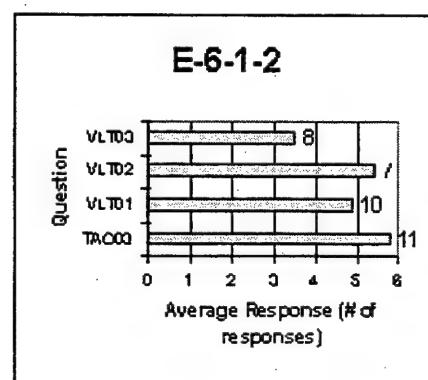
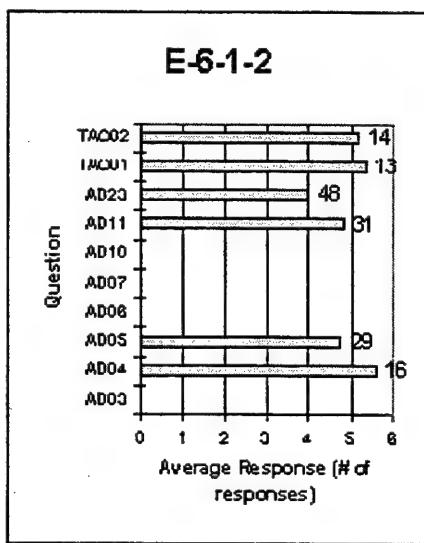
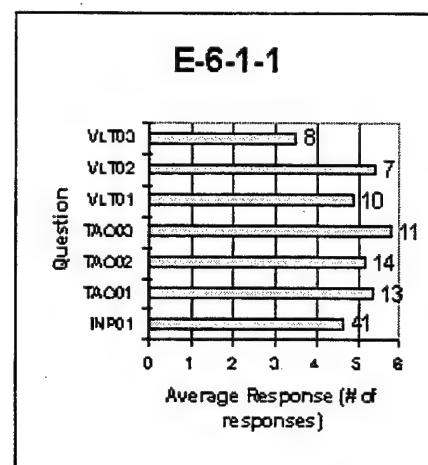
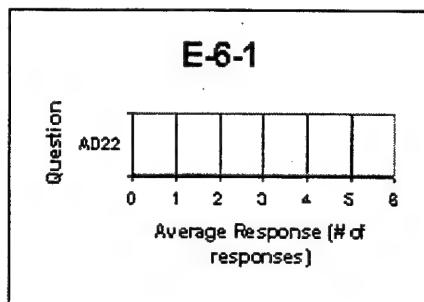


Figure 11. Sample Page of Summary Charts

This unusual coding strategy is followed due to a somewhat obscure, but hard limitation in Excel 97 in the number of charts that can be created within one application. The limiting factor is actually a maximum number of 256 fonts (including character size) that can be used by Excel 97 at one time. If this number is exceeded, the result is a “Not enough memory” error message and no more charts are produced. This is an Excel 97 bug that is documented by Microsoft, and is not necessarily a limitation in other versions of Excel [Ref. 14]. Because each of the charts created in this application uses multiple fonts, Excel is unable to produce more than 28 charts at a time.

The solution implemented to overcome this limitation is to open multiple applications of Excel, each of which is kept under the 256-font ceiling by limiting the number of charts produced to 24. This is coded so the process is almost transparent to the user, as only one application of Excel becomes visible at a time in the print preview mode. As the user views and closes or prints each four-page section, that application is closed as the next application is made visible. The exception is the last application, which as mentioned, remains open for reference. In the extreme case of perhaps hundreds of charts, this approach could have difficulties as many more instances of Excel are opened, putting a strain on system resources. However, it is not anticipated that the charting requirements for OPEVAL will change significantly from those generated by the OT-IID data this implementation was tested on. And presumably this problem will be fixed in future versions of Excel so a more robust solution may be implemented.

V. OT-IID DATA ANALYSIS

The second main effort of this thesis is to analyze the data gathered during the last operational assessment, OT-IID, with the goal of obtaining insights that will enable a more efficient and reliable data collection effort in subsequent test periods, particularly OPEVAL. Two distinct subsets of data are examined; V-22 pilot responses and responses collected via the maintenance questionnaire. The first is explored using basic summary statistics, in an effort to isolate trends and biases in pilot responses. The second is analyzed using a formal cluster analysis, in order to identify ways in which the maintenance questionnaire may be made shorter and more concise. Both analyses illustrate analytical methods that can be used to gain insights on the test data and provide means by which to make the data collection effort more efficient.

A. PILOT DEMOGRAPHICS

1. Data

At the time OT-IID was conducted, the MOTT had only six pilots qualified to fly the V-22. Five potentially relevant aspects of the pilots' individual backgrounds were provided by the MOTT, and are summarized in Table 9. Their backgrounds cover three

Pilot	Rank	Community	Service	Flight Hours*	WTI	Responses**
A	Maj	CH-53E	USMC	2647	Yes	955
B	Maj	CH-46E	USMC	2297	Yes	830
C	LtCol	MH-53J	USAF	4179	No	1189
D	LtCol	CH-53E	USMC	3848	Yes	1005
E	Major	CH-46E	USMC	2503	Yes	722
F	Major	CH-46E	USMC	3359	No	602

*As of May 1999

**OT-IID, does not include responses of NA

Table 9. MOTT Pilot Demographics

aircraft communities, two services, two ranks, and individual accumulated flight times from approximately 2300 to 4200 hours. Some are graduates of the Weapons and Tactics Instructor Course (WTI), the Marine Corps' advanced tactical training course for aviators, and some are not. Over the two-month operational assessment, these six pilots together answered almost 6400 questions distributed over dozens of post-event questionnaires. [Ref. 9]

In order to simplify the analysis in a logical and intuitive manner, each mission area in VQT, and therefore all questions assigned to it, is assigned a category number corresponding to one of six broad evaluated areas. These categories and the assignment of mission areas to them are shown in Table 10. The categories are strictly an ad hoc

Category	Description	Mission Areas
1	Mission Operations	Self Deployment FARP Mission Planning AIE/Aerial Delivery Escort Assault Support Operations
2	General Flight Operations	VLATT FAMS CALS/MALS TAFDS FCLP NATOPS
3	Communications and Navigation Systems	Comm systems (secure) Navigation Comm Systems (long-range)
4	Defensive Capabilities and Systems	Threat Density Spatial Coverage Threat Detection CM Program Selection CM Selected CM Programming
5	Safety	Ground Safety Human Factors Flight Safety
6	Night Operations	Night Vision Device

Table 10. Mission Area Category Assignments

structure based on the author's knowledge and operational experience as an aviator, and correspond to six topics into which the evaluated mission areas could reasonably be aggregated.

All responses provided by the pilots are extracted from the VQT database with a simple query and joined with the identity and characteristics of the evaluator. Responses of "NA" are not included. In addition, responses to the single question under the "Overall Comments" mission area are excluded due to the non-specific nature of the question. The resulting table of data consists of 5303 numerical responses to individual questions, each response being matched to the evaluator, his particular demographic "variables", the question asked, the mission area and category it addressed, and the questionnaire from which it came. The breakdown of the number of questions answered by each pilot is provided in Table 9.

2. Method of Analysis

For a first look at the data, the table is imported into S-Plus 4.5 and a histogram is created for each pilot of all responses provided over all categories (Figure 12). A cursory glance at Figure 12 reveals a wide variety of response patterns. All distributions show a negative skew to some degree, with a distinct mode at either five or six. The differences, however, are fairly striking.

At one extreme is pilot D, whose overwhelming majority of responses are sixes. A look at the numbers reveals that out of 1005 total numerical responses, this pilot responded with a score of four only five times during OT-IID, and never responded with anything lower. At the other end of the spectrum is pilot E, who exhibits a bimodal

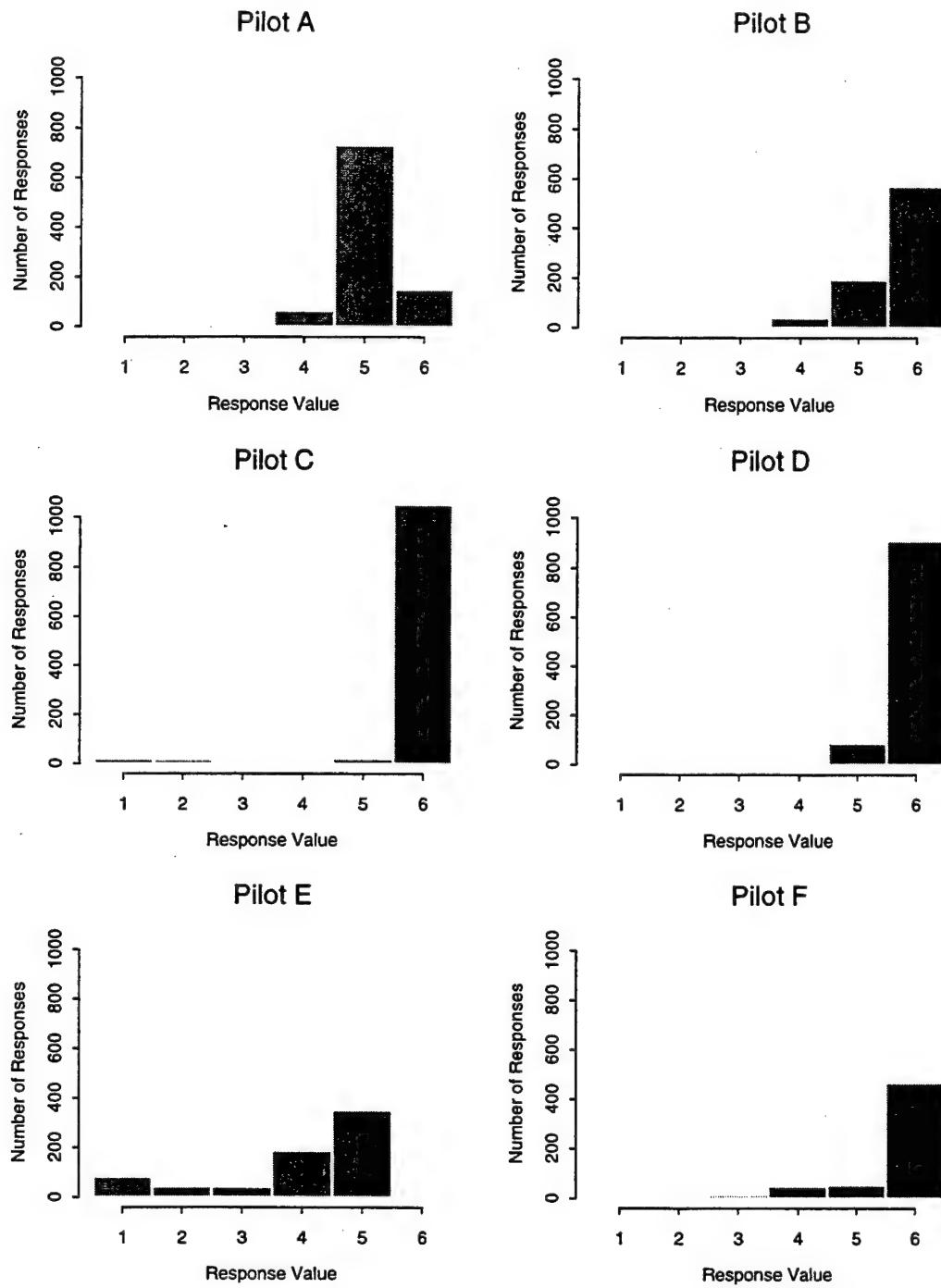


Figure 12. Pilot Histograms by Response Value

distribution of responses and is obviously much more inclined to respond with lower number than any of the other MOTT pilots. Out of 722 numerical responses given by pilot E, only five of them are sixes, while 174 responses of three or less are recorded.

The difference between these two pilots is substantial, and certainly bears further scrutiny. Pilot D, with his overwhelming tendency to respond in the positive end of the spectrum, could potentially do the test and evaluation process an injustice by not providing a sufficient level of discrimination. Pilot E, on the other hand, may be holding the aircraft to a standard that is inflated and possibly unrealistic, which could lead to skewed data and inconsistencies in the responses collected from the pilots.

The boxplots of Figure 13 provide an alternate look at the same data, this time each pilot's data being divided into the six categories of questions. Because of the integer nature of the data, the data for some of the pilots resulted in interquartile ranges of zero. Thus all points but outliers are plotted on a single line. Pilots A, C, and D show very little variability in their response patterns, while pilots B, E, and F are somewhat more varied in their opinions.

While it certainly stands out from the others, the response pattern of pilot E is not in itself highly unusual. It may be demonstrating what could arguably be considered a healthy and possibly warranted distribution of opinions concerning the merits of the V-22 in terms of the six categories presented. What is unusual is that pilot E is the only one of the six who exhibits this type of response pattern. With regards to all subjects except category four, concerning defensive capabilities and systems, his skepticism is not shared by the other pilots. Even in the case of category four his responses are comparatively

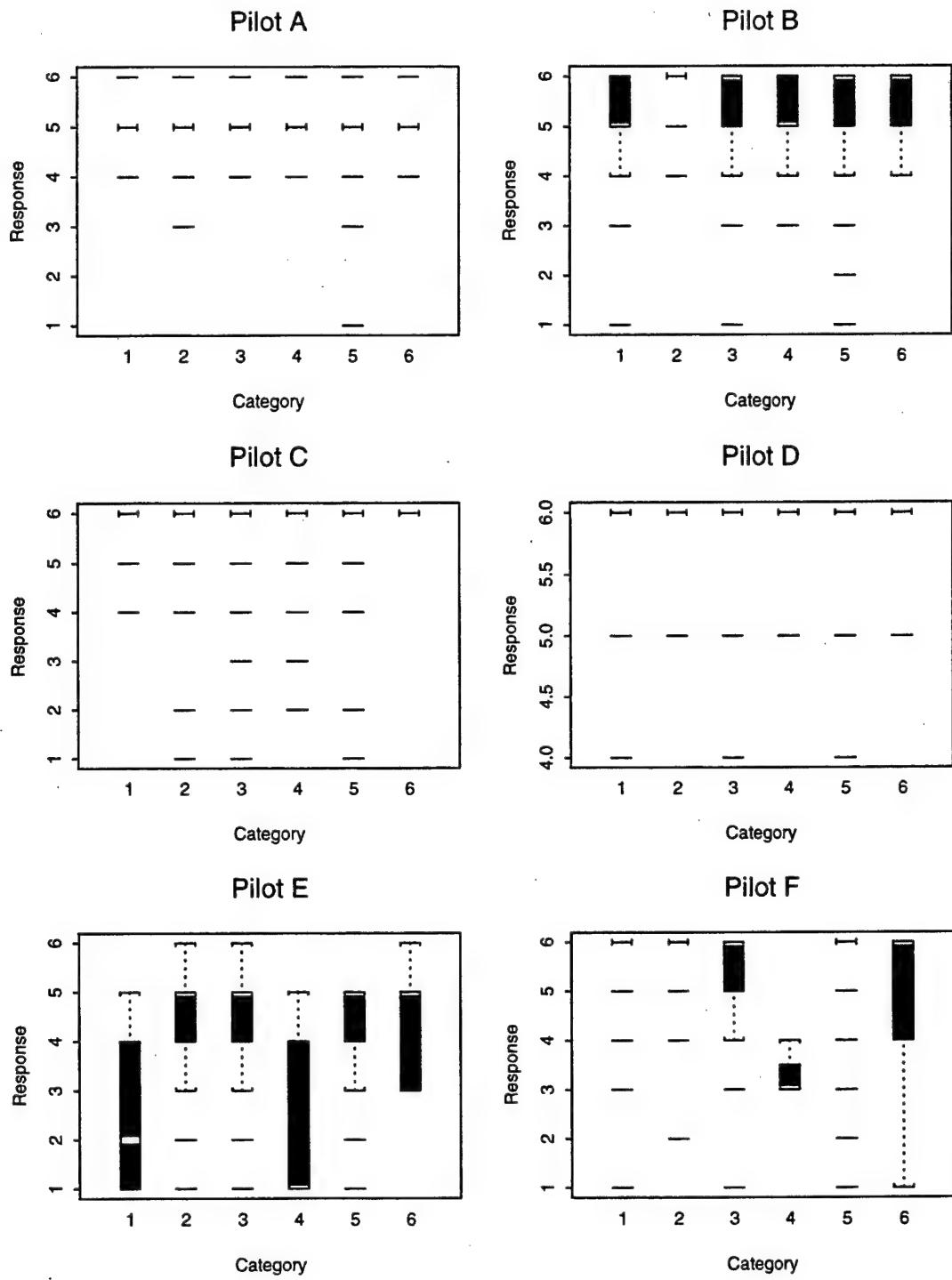


Figure 13. Pilot Response Trends by Category

extreme to the negative side. Looking at the group as a whole, pilots B and F are exhibiting what could be considered the most "typical" response patterns. That is, their responses demonstrate a pattern of discrimination in terms of what aspect of the aircraft is being evaluated, but not nearly the negative response tendencies of pilot E.

As a simple means of investigating any trends that may be occurring, the data table is imported into an Excel 97 spreadsheet where it is further manipulated using Excel pivot tables. The use of pivot tables facilitates quick and easy manipulation of the data into various contingency tables, whose changes can be immediately reflected on a chart. Figure 14 is the first of the charts, which plots the average response from each pilot, broken down by the six question categories.

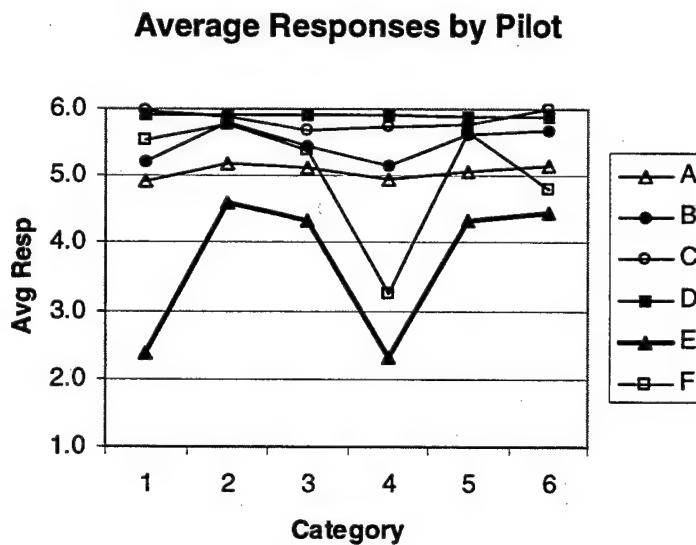


Figure 14. Average Responses by Pilot for All Categories

One trend that was noted in the previous two graphs is again apparent in this chart, in that pilot E's scores are consistently and significantly lower than all others across every category. It is apparent at this point that regardless of how the data are

grouped, whatever group pilot E belongs to will tend toward lower relative scores due to his relatively extreme response pattern. Because the purpose of this analysis is to discover latent biases occurring in pilot responses due to demographic factors, pilot E's response pattern becomes a confounding factor.

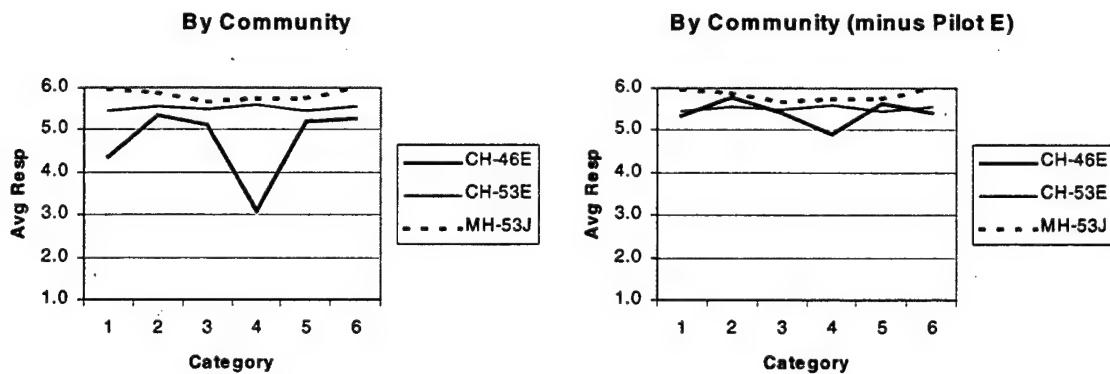


Figure 15. Average Responses by Aircraft Community

Figure 15 is provided as an example of what is typical of the various groupings of this data. The chart on the left includes responses from all pilots, while the other represents the same data minus pilot E's responses. By examining only the chart on the left, it cannot be determined whether the relatively low trend of responses from CH-46 pilots is attributable to some general bias within that aircraft community or whether it is merely a reflection of pilot E's presence in this group. The second chart clearly illustrates that the presence of pilot E in that group had a pronounced effect, since without his data the CH-46 line no longer presents a consistent trend. For this reason, pilot E was removed from the data set and subsequent comparisons made with the resulting abbreviated table.

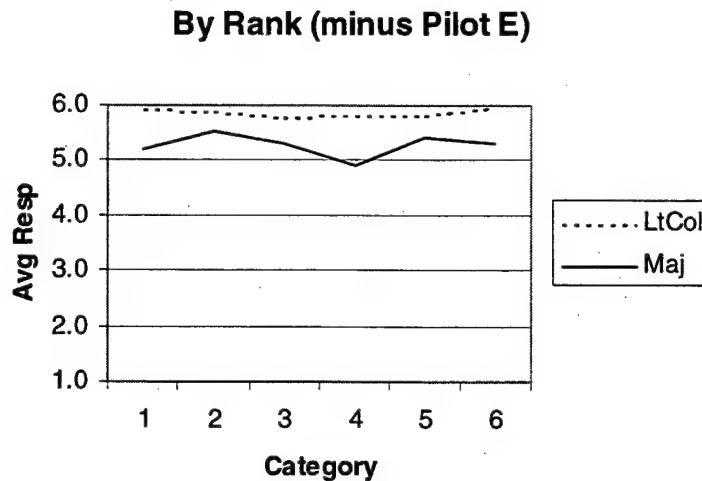


Figure 16. Average Responses by Pilot Rank

Figure 16 presents a comparison of average responses by the ranks of the individual pilots. Of the five remaining pilots in the data table, three are majors and two are lieutenant colonels. The upper line represents the scores of the lieutenant colonels, the lower being those of the majors. There does appear to be a sizeable difference across all categories, with the lieutenant colonels exhibiting a greater tendency to respond with higher scores in all categories. In categories one and four that difference appears to be almost one.

The next plot, Figure 17, which is very similar to the previous plot, presents the data broken down by the accumulated flight hours of each pilot at the end of OT-IID. An arbitrary figure of 3000 hours is used as a discriminator, since that number separates the six pilots into two groups of three. For the purposes of this study, above 3000 hours a pilot is rated "High", while below that he is rated "Low". One would expect this variable to be highly correlated with the ranks, and the similarity of the charts appears to confirm

By Flight Hours (minus Pilot E)

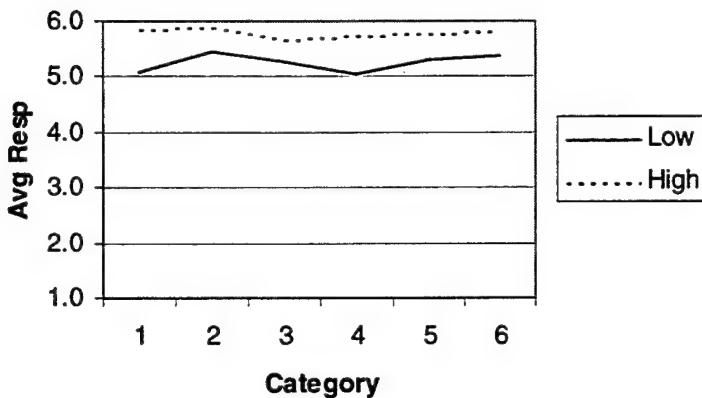


Figure 17. Average Response by Flight Hours (Low is < 3000 hrs.)

that suspicion. In general, an individual who is higher in rank will have accumulated more flight hours than his subordinates. Thus nearly the same pattern of responses is generated as with the comparison by rank. It is not exactly so in this case because one of the "high" flight-time pilots is a major.

Figure 18 represents the data categorized by pilots who have and have not attended the WTI course or an equivalent. The differences depicted between groups become much smaller than on previous charts. The exception is in category one, regarding mission operations, where the WTI graduates tend to respond significantly lower than the non-WTI graduates. This is not a surprising trend, since the WTI course is extremely mission oriented and produces the tactics instructors for all Marine Corps aviation communities. These individuals could be reasonably expected to evaluate mission-specific issues with perhaps a more critical eye than their non-WTI peers.

By WTI Graduate (minus Pilot E)

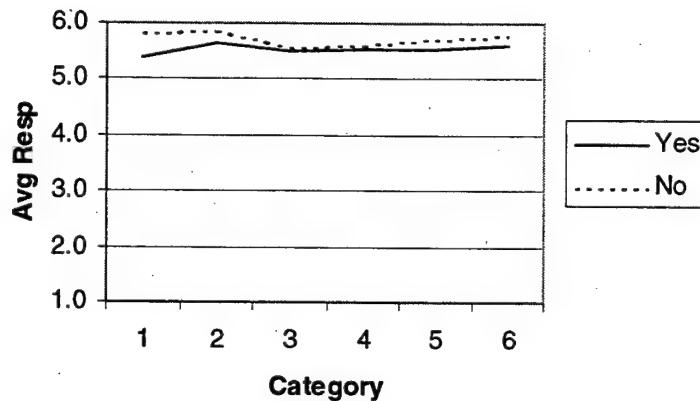


Figure 18. Average Response by WTI Graduate (Yes = graduate)

The last comparison made is between services, as illustrated in Figure 19. The Air Force is represented by the top line, the Marine Corps being shown by the lower. While there appears to be a consistent, although not alarming difference between the two

By Service (minus Pilot E)

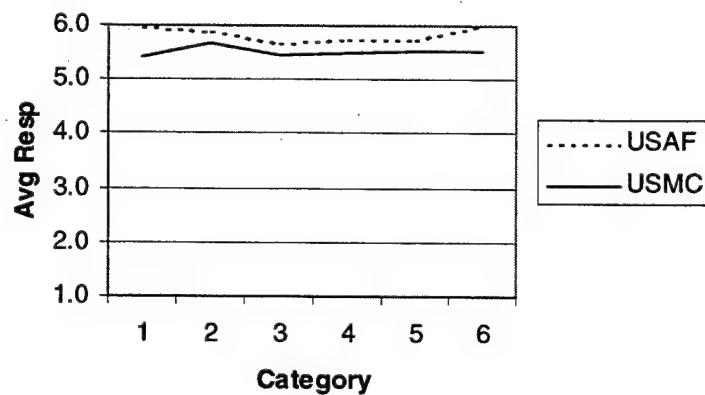


Figure 19. Average Response by Service

lines, this chart is of questionable value. This is due to the fact that only one of the pilots participating in OT-IID is a member of the Air Force, while the remaining five are

Marines. Although he does exhibit a consistently higher expected response across all categories than the Marines', it cannot be claimed with a sample size of one that this individual is truly representative of the Air Force. His difference may be due to a personal trait, it may be a product of the aircraft community from which he came, or it may be due to something else altogether. Whatever the reason, the value added to the analysis by this chart is minimal, beyond the information gained by plotting this individual pilot's response pattern relative to the other pilots.

Any trend of interest may be further investigated with the use of Excel pivot tables. For example, as Figure 15 shows, category four questions appear to produce significantly lower responses from CH-46 pilots than from pilots of the other communities. A pivot table is used to create a summary of category four questions by pilot, the results of which are shown in Table 11. Examination of the data in this form reveals that the average for the CH-46 community is being depressed in part by the low inputs from a single individual, pilot F, who has actually only provided four responses to two questions out of the eleven in this category. Three of his four ratings are in response to the question that already has the lowest average response from the other pilots. Without the benefit of other, perhaps higher responses to the remaining questions, his low ratings will tend to bias the CH-46 average low. Any trends or anomalies seen in the data must of course be investigated and judged by the analysts and decision-makers of the MOTT as to their accuracy and relevancy, and whether any sort of corrective action is appropriate. But as this example illustrates, the use of a pivot table facilitates the identification of outliers or trends that may require further investigation.

Question number	Data	Pilot				
		A	B	C	D	F
193	Average	5.33	6.00	6.00	6.00	
	Count	6	3	5	8	
198	Average	5.40	6.00	6.00	6.00	
	Count	5	1	5	8	
199	Average	4.00		6.00	5.86	
	Count	1		5	7	
200	Average	4.67	5.33	6.00	5.88	
	Count	3	3	5	8	
201	Average	4.00	3.80	4.25	5.50	3.00
	Count	2	5	8	10	3
204	Average	5.00	6.00	6.00	6.00	
	Count	3	2	5	7	
205	Average	4.67	5.50	6.00	6.00	
	Count	3	4	5	8	
206	Average	4.50		6.00	6.00	
	Count	2		5	1	
207	Average	5.00	5.50	6.00	5.86	
	Count	2	2	5	7	
208	Average	5.00	5.00	5.83	6.00	4.00
	Count	5	4	6	8	1
209	Average	5.00	5.00	6.00	6.00	
	Count	1	1	5	7	

Table 11. Pivot Table Report of Category Four Pilot Responses

3. Results

Several comments must be made concerning the significance of the observed differences in these plots. A first pass at this data using analysis of variance (ANOVA) to formally assess the significance in differences between any of these groups demonstrates the difficulty in formulating an appropriate test for this data. As Figures 12 and 13 would seem to indicate, an assumption of equal variances is not justified, and when several ANOVA tests are attempted on the data this is confirmed by the distinct heteroscedasticity of the residuals in every case. In addition, without exception the quantile-normal plots of the residuals in each test result in highly non-linear plots, indicating nonnormality of errors. In light of these facts, there seems little justification

for any assumption of normality, equal variances, or even of fitting a distribution to model the overall pattern of responses for the pilots.

When the nature of the data is examined, an assumption of independence cannot even be made between responses or variables. All responses were provided by only six individuals who, in many cases, answered the same question on multiple occasions. This creates a high level of dependence between cases. There is also a high amount of confounding between some of the demographic variables. For example, a higher-ranking individual tends to have accumulated a higher number of flight hours, and the sole representative of the Air Force in this test is also the sole representative of the MH-53J community. Thus, even a nonparametric significance test such as the Mann-Whitney or Friedman Test is difficult to justify, given the amount of dependence within the data.

The trends noted must still be judged, however, in terms of operational significance, as distinct from statistical significance. Although a formal test of statistical significance may be capable of detecting a very small departure from a given null hypothesis, particularly since this data set is rather large, that departure may not be significant in practical, or operational terms. The extent to which an observed difference has the potential to affect the outcome of the event must be the factor that determines the level of concern that is warranted. For example, the tendency for WTI graduates to respond in a more critical manner to category one questions, as illustrated in Figure 18, may be distinctly noticeable on the chart and highly significant in statistical terms. But in practical terms, it may have no bearing on the outcome of the operational test. Devore [Ref. 15:p. 340] provides a formal discussion of statistical versus practical significance.

Even in lieu of formal statistical tests for significance, several noteworthy trends can be extracted from the charts. The trend for pilot E to respond much lower than the other pilots is arguably the most important of the patterns observed, and is an issue that must be resolved for future test events. Given the contrast between this pilot and pilot D, whose tendency is to respond with only high numbers, there may be a problem, for example, with standardization among pilots in terms of aircraft performance expectations, grading criteria, or possibly interpretation of the scale response alternatives or even the questions themselves.

In terms of demographic factors, the tendency for the more senior, higher flight-hour pilots to respond with higher ratings across all categories is certainly noteworthy, and appears to be the most pronounced trend produced by the factors examined. A comparison of categories among the charts reveals that there appears to be a relatively high degree of consensus among the pilots with regards to categories two, three, and five, the topics of General Flight Operations, Communications and Navigation, and Safety. Categories one, four, and six, corresponding to Mission Operations, Defensive Capabilities and Systems, and Night Operations, bring out more variability in responses. The profile of the pilots who tended to respond with the lowest ratings in these three categories is a major from the CH-46E community with less than 3000 flight hours, who has graduated from WTI. On the other extreme, the most positive responses were provided by pilots fitting the profile of a lieutenant colonel from the CH-53E or MH-53J community, who has more than 3000 hours and is not a WTI graduate.

B. MAINTENANCE QUESTIONNAIRE

1. Data

The maintenance questionnaire used for OT-IID contains seventeen questions, and remained the same throughout the test period. These questions are presented in Table 12. By the end of the two-month test period, 585 maintenance questionnaires had been administered and entered into VQT. These questionnaires were completed by thirty-seven individuals, and covered a wide variety of maintenance actions over the two month test period. The goal was to administer a questionnaire for every maintenance

1. Rate the maintainability of the V-22 in field environments with limited support services.
2. Rate the overall ease of maintenance on the V-22.
3. Rate the acceptability of IETMs fault isolation procedures for troubleshooting.
4. Rate the suitability of the BIT/AMEGS output as a trouble shooting aid for this task.
5. Rate the suitability of the brick for uploading, storing, and downloading of data between the V-22 and AMEGS.
6. Rate the suitability of the maintenance facilities to work on this task.
7. Rate the acceptability of the V-22 to be compatible with existing servicing/support systems.
8. Rate the effectiveness of the training you received to perform this task.
9. Rate the suitability of the Packaging, Handling, Shipping/Storage & Transportation of aircraft spares.
10. Rate the suitability of the number of personnel (by MOS) required by IETMs to accomplish this task.
11. Rate the accessibility to the system being maintained for this task.
12. Rate the acceptability of the fasteners.
13. Rate the suitability of the tools and support equipment to accomplish this task.
14. Rate the suitability of the IETM to state clearly all information required to perform this task.
15. Rate the accuracy of the IETMs for this task.
16. Rate the overall safety of this task.
17. Please take this opportunity to describe your overall assessment of this event. BE SPECIFIC.

Table 12. Original Maintenance Questionnaire

action accomplished, with an actual return rate of approximately 95% by the end of OT-IID [Ref. 9]. For reasons discussed in Chapter II, every questionnaire collected contains at least one question that was not applicable to the task accomplished, and therefore has

at least one response of "NA" in the database. Some questions actually are answered "NA" in the majority of cases.

As the responses are stored one to a row in the VQT "ANSWERS" table, a crosstab query in Access is necessary to extract the data such that all responses are assigned by row to the questionnaire from which they were collected. A crosstab query is the Access equivalent of an Excel pivot table, and performs in exactly the same manner. The resulting table is 585 rows of responses, each row corresponding to a questionnaire, and 17 columns corresponding to the questions, in this case to be treated as the variables. Because the final question solicits comments only, it is not relevant to this analysis and is dropped from the table, leaving 16 columns of data.

2. Method of Analysis

A sensible approach to optimizing the maintenance questionnaire is to eliminate redundancy between questions, and attempt to gather the same information using a fewer number of questions. Cluster analysis is an analytical method for accomplishing this using formal statistical tools. The objective of conducting cluster analysis on this data is to group questions in such a way that those belonging to the same cluster resemble each other, but are dissimilar from those in other clusters. A complete discussion of cluster analysis is provided by Everitt [Ref. 16]. In this case, the similarity or dissimilarity of questions is measured in terms of the responses elicited. It is hoped that these groupings, or clusters, will provide a logically sound basis for rewording as a single, more concise question, and result in a maintenance questionnaire consisting of fewer questions.

The “agnes” and “diana” functions of S-Plus 4.5 are employed to approach this problem from opposite directions. Both functions use hierarchical clustering algorithms, “agnes” using an agglomerative hierarchical clustering method and “diana” using a divisive hierarchical clustering method [Ref. 17:p. 501]. Hierarchical clustering algorithms create a complete hierarchy of clusterings for a given data set of n objects, from n individual clusters at one extreme to one cluster containing all objects at the other. An agglomerative method begins with each of n objects forming its own cluster and one by one merges objects and clusters until all objects are contained in one cluster. This is accomplished by successively joining the clusters that have the smallest dissimilarity between them. A divisive method, on the other hand, starts with all objects in one cluster, and successively splits clusters until all objects form their own cluster. The splits in this case are based on the largest dissimilarities between any two possible clusters that may be created.

Two metrics are used in this analysis to measure the dissimilarity. A matrix based on the “euclidean” metric uses euclidean distance, or root sum of squares, as a measure. The “manhattan” metric, on the other hand, uses the sum of absolute differences between responses. The method used by the “agnes” function to measure distances between clusters is also selectable by the user. For this analysis the default setting of “average” was used, which averages the distances between the points in one cluster and the points in the other cluster.

Figure 20 is a pair of dendograms that illustrate the output of the “agnes” and “diana” functions on the maintenance data, using a “euclidean” distance metric. A

dendogram graphically portrays to the user which objects are joined into clusters at what point in the process, and the distance values associated with those clusters. Comparison of the two dendograms reveals that the output of the agglomerative and divisive methods

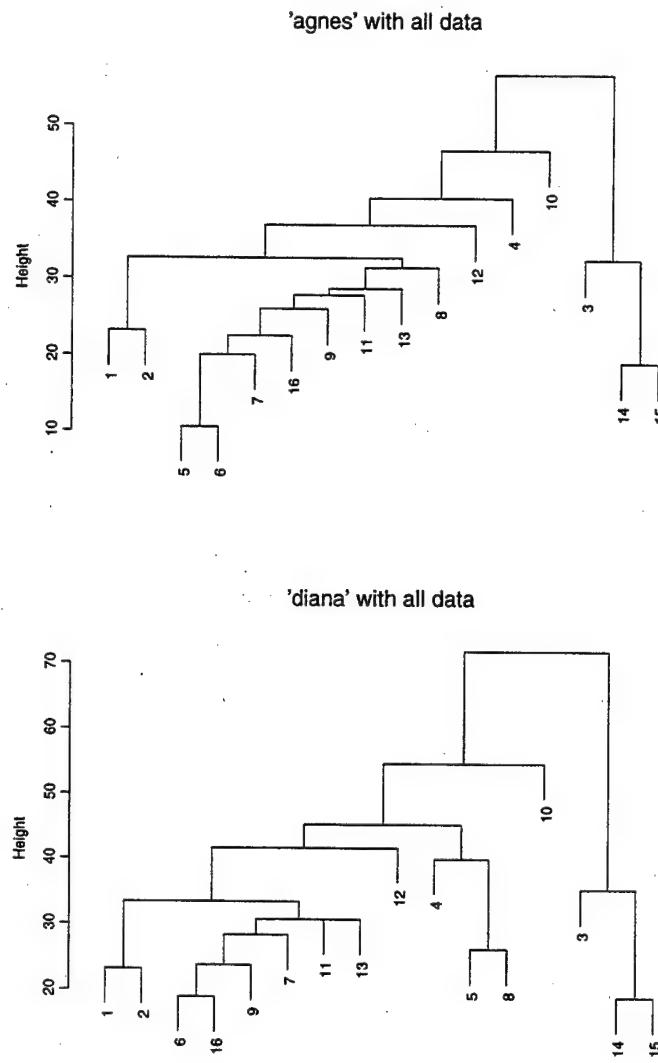


Figure 20. Cluster Analysis Plots on Complete Maintenance Data Set

is very similar. As the clusters at the bottom of the charts are those that are most similar, they correspond to the questions that are of the most interest in this analysis. They

provide most nearly the same pattern of responses, and are thus candidates for possible integration into a single, perhaps rephrased question. This is of course dependent on the exact wording of the questions, and whether integration makes sense in the context of the specific information sought by the individual questions.

Examination of the actual questions that have been grouped on the dendograms reveals that most of the initial clustering does, in fact, make sense in practical terms. For instance, questions one and two, comprising the fifth and third cluster created by "agnes" and "diana", respectively, address very highly correlated issues. If the overall ease of maintenance on the V-22 is high, then the maintainability in field environments with limited services would naturally be rated similarly. These two questions are therefore prime candidates for rewording and collapsing into a single question. However, not all groupings depicted are this logical. For example, question five is clustered by "agnes" with question six, and by "diana" with question eight, both clusters occurring at or near the bottom of the hierarchy where they are among the most significant. In terms of subject matter however, question five does not bear any obvious relationship to either of the questions with which it was paired, and would seem to belong much farther up in the hierarchy where the dissimilarities are much greater.

A closer examination of the data reveals that, while 12 of the 16 questions have over 300 numerical (non-NA) responses, 10 of those having over 400, question five has been answered "NA" all but 45 times, the lowest response rate of all questions. In addition, as Figure 21 illustrates, a histogram of the overall response pattern shows that the distribution of responses leans heavily towards responses of fives and sixes. The

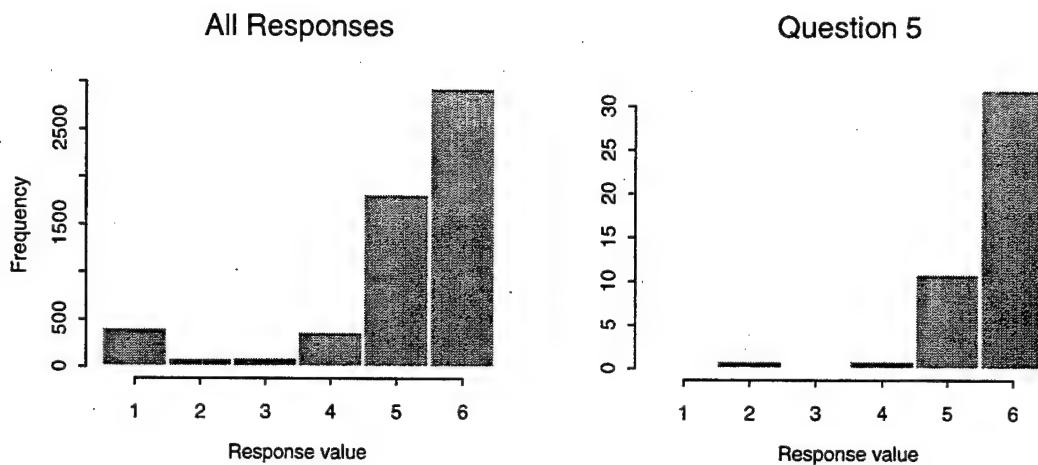


Figure 21. Maintenance Questionnaire Response Patterns

histogram of the responses for question five shows the same trend. In light of these two facts, the possibility is greater for this question than others that it could exhibit a response pattern that is just by chance very similar to other, totally unrelated questions.

Question five is therefore removed from the data table and the same analysis conducted on the remaining data. Because the numbering of the questions is then altered,

- 1. Rate the maintainability of the V-22 in field environments with limited support services.
- 2. Rate the overall ease of maintenance on the V-22.
- 3. Rate the acceptability of IETMs fault isolation procedures for troubleshooting.
- 4. Rate the suitability of the BIT/AMEGS output as a trouble shooting aid for this task.
- 5. Rate the suitability of the maintenance facilities to work on this task.
- 6. Rate the acceptability of the V-22 to be compatible with existing servicing/support systems.
- 7. Rate the effectiveness of the training you received to perform this task.
- 8. Rate the suitability of the Packaging, Handling, Shipping/Storage & Transportation of aircraft spares.
- 9. Rate the suitability of the number of personnel (by MOS) required by IETMs to accomplish this task.
- 10. Rate the accessibility to the system being maintained for this task.
- 11. Rate the acceptability of the fasteners.
- 12. Rate the suitability of the tools and support equipment to accomplish this task.
- 13. Rate the suitability of the IETM to state clearly all information required to perform this task.
- 14. Rate the accuracy of the IETMs for this task.
- 15. Rate the overall safety of this task.

Table 13. Numbering Scheme for Abbreviated Questionnaire (Questions 5 and 17 removed)

Table 13 is provided for reference with the new numbering scheme. As the dendograms of Figure 22 illustrate, the resulting clustering schemes for the two methods exhibit a few minor differences from the initial analysis with the full data set, but for the most part

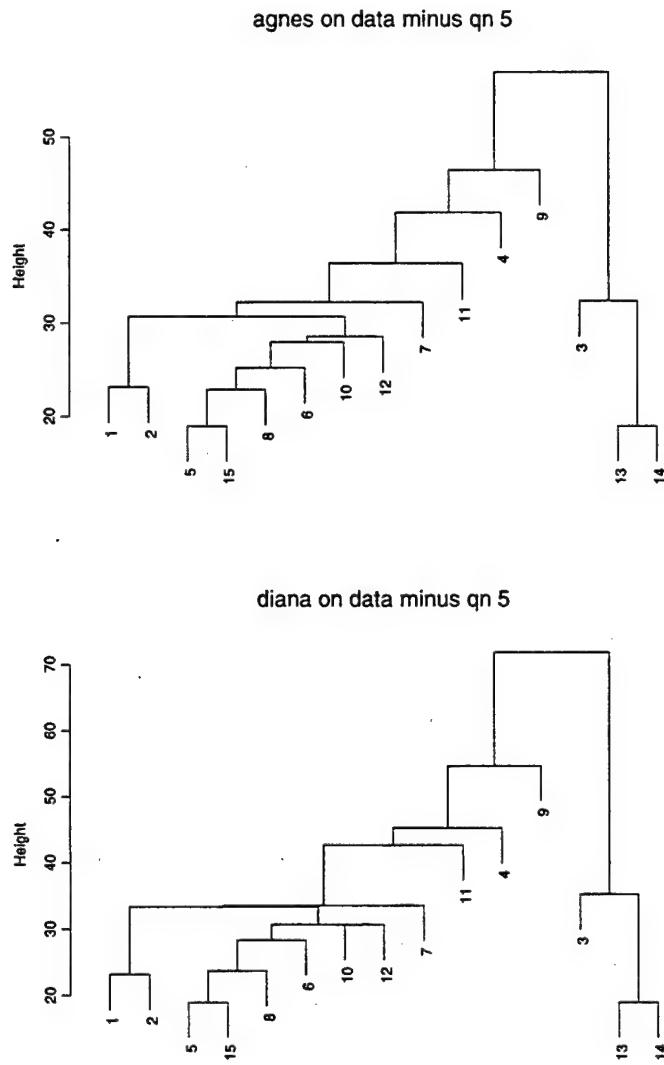


Figure 22. Cluster Analysis Results on Abbreviated Data Set

remain the same. One striking and reassuring feature that presents itself here is that with question five removed, the clustering schemes of the agglomerative and divisive methods

are now in total agreement, although there are slight differences in the order in which the clusters occur. This would seem to validate the results to some degree.

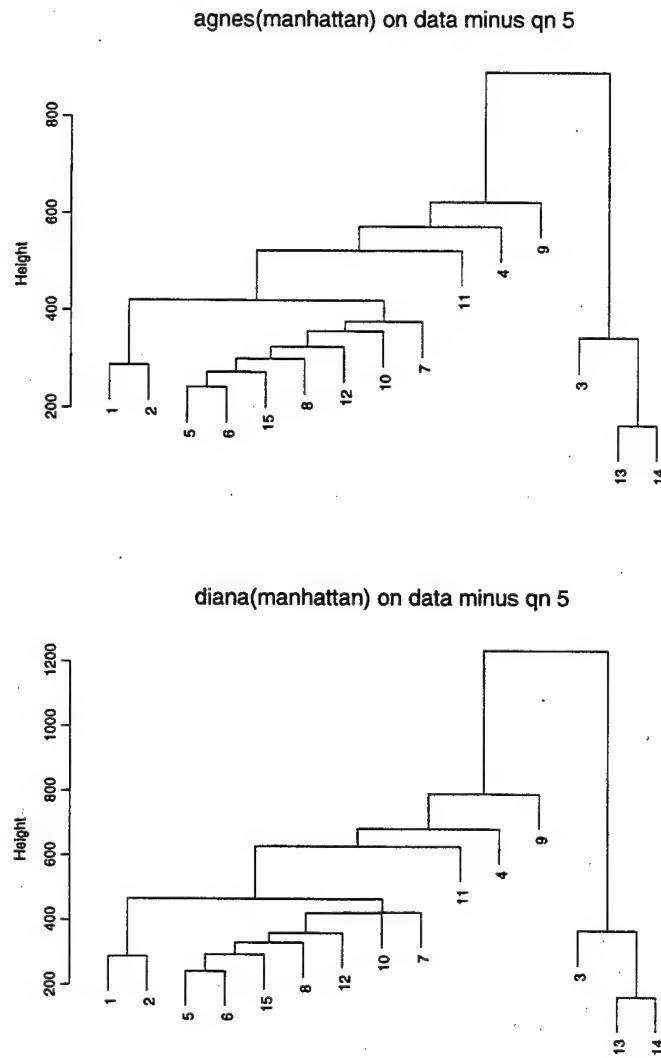


Figure 23. Cluster Analysis Results Using "Manhattan" Distance

As a second informal means of cross-validation, another analysis on the smaller data set is run with the metric for computing dissimilarities set to "manhattan." Figure 23 illustrates the resulting dendograms for "agnes" and "diana". The results change very

little for either method from those attained through the "euclidean" distance measure. As before, the clustering schemes for the two methods are nearly completely in agreeance. Questions five, six, fifteen, and eight are clustered in slightly different orders than in the preceding analysis, but in terms of this study the change is of little significance.

3. Results

When the questions themselves are examined in terms of the clusters that have been produced, several opportunities for aggregation of questions become apparent.

Table 14 is an example of how several of the questions could be collapsed using the previous clustering scheme as a guideline. The questionnaire that previously contained

New	Old	
1.	1, 2	Rate the maintainability of the V-22 in field environments with limited support services.
2.	3	Rate the acceptability of IETMs fault isolation procedures for troubleshooting.
3.	4	Rate the suitability of the BIT/AMEGS output as a trouble shooting aid for this task.
4.	5, 6, 15	Rate your ability to safely complete this task in terms of maintenance facilities and existing servicing/support systems.
5.	7	Rate the effectiveness of the training you received to perform this task.
6.	8	Rate the suitability of the Packaging, Handling, Shipping/Storage & Transportation of aircraft spares.
7.	9	Rate the suitability of the number of personnel (by MOS) required by IETMs to accomplish this task.
8.	10	Rate the accessibility to the system being maintained for this task.
9.	11	Rate the acceptability of the fasteners.
10.	12	Rate the suitability of the tools and support equipment to accomplish this task.
11.	13, 14	Rate the ability of the IETM to effectively present all information required to perform this task.
12.		Rate the suitability of the brick for uploading, storing, and downloading of data between the V-22 and AMEGS.
13.		Please take this opportunity to describe your overall assessment of this event. BE SPECIFIC.

Table 14. Reworked Maintenance Questionnaire (Old numbers refer to Table 13)

seventeen questions now has been reduced to twelve, using this suggested scheme. The practicality of this shortened questionnaire and the potential for any further aggregation must of course be determined by the analysts, OTDs, and maintainers of the MOTT in light of their specific data collection requirements. But cluster analysis has provided

some numerical insight into possibly more optimal arrangements of the questions, which may not have been otherwise apparent.

VI. CONCLUSIONS

A. DISCUSSION

The purpose of this thesis was to address several shortcomings in the current method the MOTT uses to collect questionnaire data. The products of this thesis are several proposed methods to assist in making the collection of operational test questionnaire data more efficient and reliable. Two software tools have been created that will directly improve the questionnaire generation and result presentation capabilities of VQT. Using data from the most recent V-22 operational assessment, two analyses were conducted that provide means to improve the maintenance questionnaire and identify potential inconsistencies in the pilot responses.

1. Questionnaire Length

A new capability was developed for VQT by which a user creating a questionnaire can automatically screen questions for inclusion on the questionnaire in terms of the adequacy of responses received to-date. This should have a direct effect on the length of questionnaires that are produced using VQT. By relating the general collection requirements to the table of data already accumulated, the data collection effort should become more efficient as the program automatically flags questions that have been sufficiently addressed. Because this process is dependent on the quantity of data already collected, the benefit observed should grow as the test period progresses. At the start of the test period, since the database is just beginning to grow, there will be no screening occurring unless the criteria are artificially manipulated by providing extreme values. But as the test progresses and the database grows, the process will increasingly

screen out questions as more of them gather sufficient responses to meet the criteria provided by the user. The result should be progressively shorter questionnaires and a necessarily increased emphasis on the questions that require more attention and evaluative effort.

In addition, a formal statistical method was demonstrated by which a repeatedly administered questionnaire may be evaluated in terms of its potential to be made smaller. Incorporating the results of a cluster analysis on the maintenance questionnaire resulted in a 25% shorter example questionnaire that should provide nearly, if not exactly, the same information as the original. Using this analytical method, the analysts could use previously collected data to make the ongoing data collection process more optimal. The degree to which aggregation of questions is practical or even desired will be dictated by the specific collection requirements of the Test Plan. In addition, the exact wording of the questions should be approved by the analysts and maintainers, themselves. By administering a smaller, more efficiently constructed questionnaire, the MOTT should benefit in terms of man-hours required to complete the questionnaire and subsequently enter the data into VQT, and the decreased capacity required to store the data.

2. Reliability of Data

A method of conducting trend analysis using summary statistics on numerical questionnaire data was demonstrated, using the responses of the MOTT pilots during OT-IID as a sample data set. The analytical method developed allows ready identification of trends or biases in the response patterns and inconsistencies in the data. For example, the analysis of numerical pilot data highlighted several distinct trends that deserve further

attention prior to OPEVAL. The first, and most prominent, is the tendency for pilot E to respond significantly lower than all other pilots across all question categories. His biased responses may lead to inconsistencies in the data gathered during OPEVAL in the form of extreme outliers. The trend identified for pilot D, on the other hand, to provide overwhelmingly positive responses, may also tend to consistently skew the data in the opposite direction. A third trend is the tendency for the higher flight-time pilots to provide higher average response values than the lower flight-time pilots across all six question categories.

All three of these trends may potentially indicate a lack of standardization among the MOTT pilots with regards to performance expectations and the scaled responses used on the questionnaires. Another possibility is that one or more pilots are responding to problems or issues that are not widely encountered or known to other pilots. All trends identified in this example should be subject to further investigation by the OTDs and analysts to determine causal factors. If left uncorrected, the result may be response patterns and inconsistencies during OPEVAL that bring into question the reliability of the data collected.

3. OTD Interpretation of Test Results

The capability of graphically displaying summaries of response data was developed for VQT. The summary charts that are now available to VQT users should be beneficial in several ways. By providing an easily comprehensible, compact summary of the questionnaire responses in graphical form, the OTDs can quickly focus on issues that deserve further attention, rather than having to manually extract all information from a

comprehensive and sometimes massive textual report. Because the charts enable the individual to easily recognize abnormal or substandard response values, this summary will help him optimize his efforts by concentrating on potential problem areas. Low average responses for a question will indicate an area that requires further investigation, perhaps into the specific commentary responses that were received for that question, or perhaps into the construction of the question itself. Low quantities of responses received will alert the OTD or analyst of a question that requires further attention in future events, and will therefore assist in the event planning process. And the summaries will undoubtedly assist the OTDs in the preparation of the final test report.

B. RECOMMENDATIONS

1. Pilot Data

a. The trends noted in this thesis should be judged by the MOTT analysts and decision-makers in terms of their potential to impact test results. Trends deemed significant should be further investigated through pilot interviews, with the objective of determining any latent causal factors or uncommon knowledge among the pilots, and implementing appropriate measures. Compensatory measures could be taken to analytically compensate for observed biases or trends, or remedial measures could be taken in an attempt to eliminate the causes of inconsistencies among pilots.

b. Standardization training for the MOTT pilots should be conducted prior to OPEVAL, with the objective of creating a common understanding and interpretation of the scaled questionnaire responses.

c. Trend analysis should be incorporated into Appendix D (Data Analysis Plan) of the Test Plan. The simple analytical methods demonstrated in this thesis can be used as a means to monitor the efficiency and reliability of the data collection effort.

2. Maintenance Questionnaire

Where appropriate, cluster analysis should be incorporated into the questionnaire development process. Using the results of the cluster analysis accomplished in this thesis, the analysts and maintainers of the MOTT should determine the extent to which specific collection requirements will permit aggregated questions. As much as practical, questions should be clustered and reworded to create a shortened maintenance questionnaire for OPEVAL.

C. FURTHER RESEARCH OPPORTUNITIES

VQT is a program that is continually changing and evolving as new requirements arise. As with any computer program, there will always exist the opportunity for an ambitious programmer to improve upon the current implementation in terms of database design, algorithm efficiency, or user interface. Similarly, the nature of this type of questionnaire data invites the application of new and inventive statistical methods, perhaps in the form of an additional capability added to VQT. Any implementation should be generic enough to apply in situations other than V-22 operational testing. The following are specific issues that arose during the course of this thesis and may be suitable for further operations research work.

1. An automated capability to extract questions with unusual response patterns, such as those with highly skewed or bimodal distributions, should be developed for VQT.

This would enable an individual to isolate questions within the database that may need further attention in terms of evaluative effort or rewording.

2. A suitable statistical method for monitoring and analyzing response patterns, and the ability to determine the significance of observed trends or differences with some practical measure should be developed for incorporation into VQT.

3. A standardized, perhaps automated method or template for question construction should be developed, in compliance with current survey research literature, that assists analysts and OTDs in consistently building optimally worded questions.

4. An automated means by which to distribute frequently asked questions in some random fashion among event participants should be created for VQT. In this manner, a particular question regarding a specific test event would be answered by perhaps only one individual, rather than several, and the average length of the questionnaires would be decreased throughout the test period.

APPENDIX A. VQT DATABASE RELATIONSHIP DISPLAYS

These illustrations portray the relationships existing between the tables and fields in VQT. Figure 24 is VQT prior to the modifications described in this thesis. Figure 25 is the normalized, less redundant version used as a basis on which to incorporate the additional screening and reporting functions.

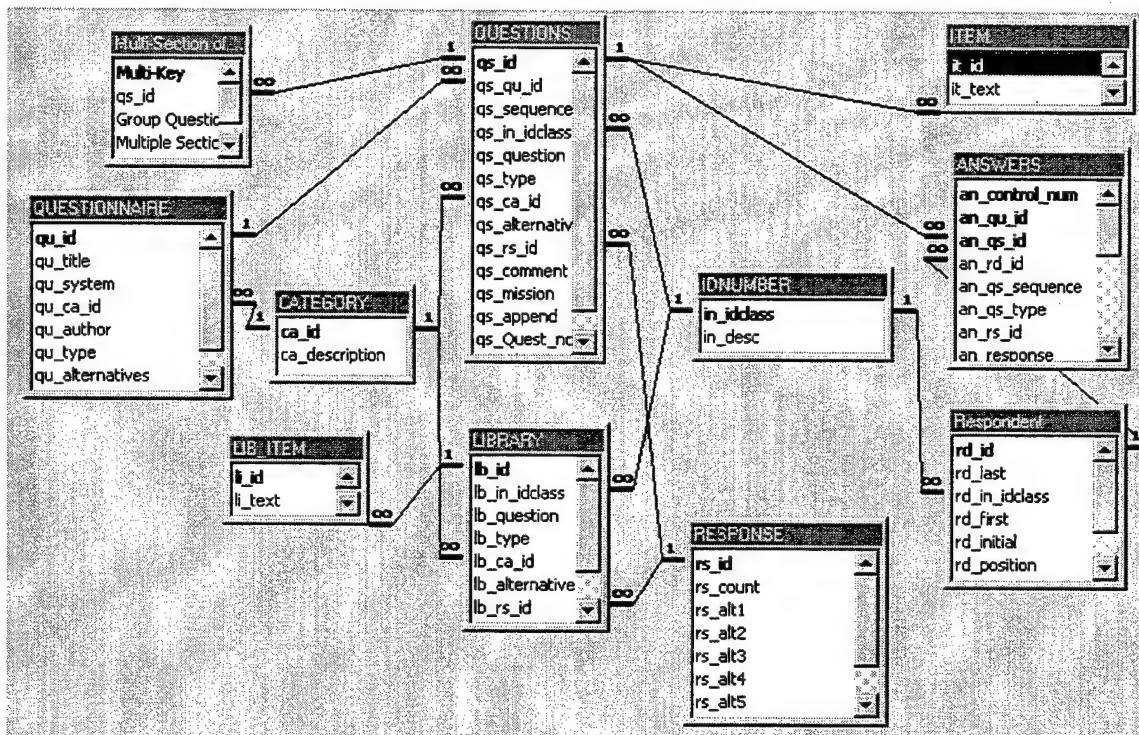


Figure 24. Original VQT Database Relationships

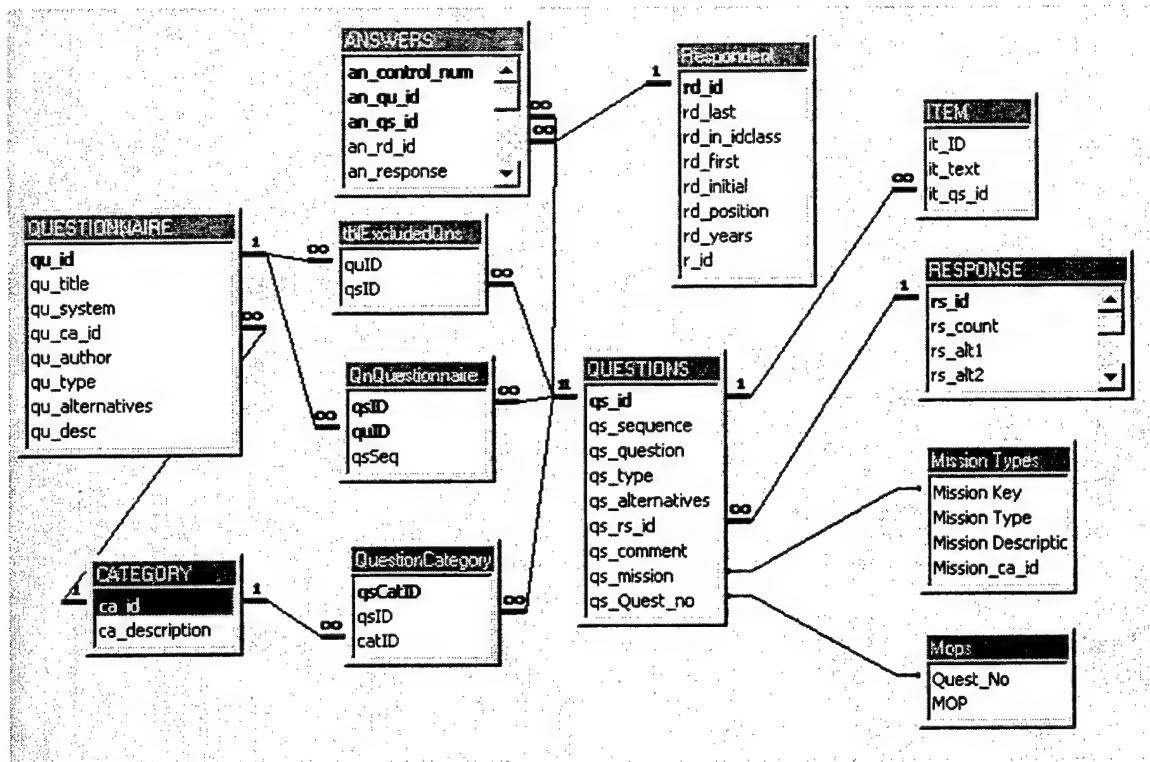


Figure 25. Revised (Normalized) VQT Database Relationships

APPENDIX B. VQT QUESTIONNAIRE SCREENING FUNCTION

The following subroutines provide the questionnaire screening function in VQT.

The first subroutine (cmdGenerate_Click()) uses criteria provided by the user to screen questions prior to inclusion on a newly generated questionnaire. The second subroutine (cmdReadd_Click()) provides the user the opportunity to over-ride the screening results, and place screened questions back onto the questionnaire prior to accepting the results.

```
Private Sub cmdGenerate_Click()

'Generate questionnaire that includes only questions that do not meet
'the user-defined criteria

Dim strSQL As String
Dim dbs As Database
Dim Temp, quID, duplic As Integer
Dim CategoryID As Integer
Dim aver, stdDev As Double
Dim varItem As Variant

DoCmd.Hourglass True
DoCmd.SetWarnings False
DoCmd.OpenQuery "qryQnStats"      'maketable query
DoCmd.SetWarnings True

duplic = Me.Text4      'read criteria from current form
aver = Me.Text0
stdDev = Me.Text2
quID = Forms!questionnaire.qu_id      'questionnaire number
CategoryID = Forms!questionnaire.Combo16      'user category

Set dbs = CurrentDb
Forms!questionnaire.Command27.SetFocus
Forms!questionnaire.Command34.Enabled = False

'Step through each mission area that user has previously selected in listbox

For Each varItem In Forms![Make Questionnaire].MissionList.ItemsSelected
    Temp = Format(Forms![Make Questionnaire].MissionList.ItemData(varItem), "0000")

    'Appends question numbers and new questionnaire number of those not meeting
    'criteria to end of QnQuestionnaire.

    strSQL = "INSERT INTO QnQuestionnaire (qsID, quID, qsSeq) " & _
    "SELECT tblQnStats.qs_ID, " & quID & ", tblQnStats.qs_sequence " & _
    "FROM tblQnStats " & _
    "WHERE (tblQnStats.qs_mission = " & Temp & ") " & _
    "AND (tblQnStats.catID = " & CategoryID & ") " & _
    "AND ((tblQnStats.Dups < " & duplic & ") " & _
    "OR (tblQnStats.Dups = Null)) " & _
    "OR ((tblQnStats.Avg < " & aver & ") " & _
    "OR (tblQnStats.Avg = Null)) " & _
```

```

"OR ((tblQnStats.SD > " & stdDev & ") " & _
"OR (tblQnStats.SD = Null)); " 

dbs.Execute (StrSQL)

'Appends question numbers and questionnaire number of those meeting
'criteria to end of tblExcludedQns.

StrSQL = "INSERT INTO tblExcludedQns (qsID, quID) " & _
"SELECT tblQnStats.qs_ID, " & quID & _
" FROM tblQnStats " & _
"WHERE (tblQnStats.qs_mission = " & Temp & ") " & _
"AND (tblQnStats.catID = " & CategoryID & ") " & _
"AND (tblQnStats.Dups >= " & duplic & ") " & _
"AND (tblQnStats.Avg >= " & aver & ") " & _
"AND (tblQnStats.SD <= " & stdDev & ");"

dbs.Execute (StrSQL)

```

Next varItem

```

DoCmd.Hourglass False
DoCmd.OpenForm "Screening Results"
DoCmd.Close acForm, "Make Questionnaire"
End Sub

Private Sub cmdReadd_Click()

'Move user-selected questions from tblExcludedQns to qnQuestionnaire
'in order to selectively disregard selection criteria

On Error GoTo Err_cmdReadd_Click

Dim StrSQL As String
Dim dbs As Database
Dim Temp, quID As Integer
Dim i As Variant

DoCmd.Hourglass True
quID = Forms!questionnaire.qu_id      'take questionnaire number from form
Set dbs = CurrentDb
Forms!questionnaire.Command27.SetFocus
Forms!questionnaire.Command34.Enabled = False

'step through all questions selected by user

For Each i In List9.ItemsSelected
    Temp = Format(List9.ItemData(i), "0000")

'Appends previously excluded question and questionnaire numbers to
'end of QnQuestionnaire.

StrSQL = "INSERT INTO QnQuestionnaire (qsID, quID, qsSeq) " & _
"SELECT " & Temp & ", " & quID & ", Questions.qs_sequence " & _
"FROM Questions " & _
"WHERE (Questions.qs_id = " & Temp & ");"

dbs.Execute (StrSQL)

'deletes those same question and questionnaire numbers from tblExcludedQns

StrSQL = "DELETE FROM tblExcludedQns " & _
"WHERE (tblExcludedQns.qsID = " & Temp & ") " & _
"AND (tblExcludedQns.quID = " & quID & ");"

dbs.Execute (StrSQL)

```

```
Next i

DoCmd.Hourglass False

List7.Requery      'refresh list
List9.Requery

Exit_CmdReadd_Click:
Exit Sub

Err_CmdReadd_Click:
MsgBox Err.Description
Resume Exit_CmdReadd_Click

End Sub
```


APPENDIX C. AUTOMATION ROUTINE FOR CHART CREATION

The following code, consisting of four subroutines, creates the summary charts in VQT. The event handler “genGraphs_Click()” is fired when the user clicks on “Summary charts” in the Reports Menu. It calls the subroutines “makeChart()” and “buildIt()”, which create each chart and format it, respectively. Finally, “newApp()” opens a new application of Excel if more than 24 charts are required.

```
Private Sub genGraphs_Click()

'Event handler for creating charts. Runs query based on desired COI and
'calculates summary statistics, opens Excel, and pastes data into Excel.
'Divides data up into COI, MOE/MOS, and MOP summaries, with max chart size
'of 10 data points. Creates charts and displays them in print preview mode.
'If more than 24 charts, opens up more applications of Excel to continue
'plotting data. Closes all but last application as user either prints or
'closes each group of 24 charts.

On Error GoTo Err_genGraphs_Click

Dim oApp As Excel.Application
Dim db As Database, qd As QueryDef, rs As Recordset
Dim i As Integer, j As Integer, k As Integer, entryCount As Integer
Dim first As Integer, last As Integer, num As Integer, remainder As Integer
Dim numChart As Integer
Dim entry As Variant
Dim chtTitle As String
Dim categoryCount As New Collection
Dim categoryName As New Collection
Dim openApps As New Collection

If [COIs] & "null" = "null" Then
    MsgBox "Please select a COI to print."
Else
    DoCmd.Hourglass True
    k = 0      'instance of an open Excel application
    Set db = CurrentDb
    Set qd = db.QueryDefs("qryGraphData")  'parameter query and parameter
    qd![Forms]![Reports Menu]![COIs] = [Forms]![Reports Menu].[COIs]
    Set rs = qd.OpenRecordset()
    SysCmd acSysCmdsetStatus, "Please wait whilst I conduct my gyrations..."
    Set oApp = CreateObject("Excel.Application")  'OLE object
    newApp openApps, rs, oApp, k
    'populate categoryName and categoryCount collections with category(COI/MOE/MOP)
```

```

'and number of questions for that category respectively
On Error Resume Next

rs.MoveFirst
entry = Trim(rs!MOP)      'category(COI/MOE/MOP)
categoryName.Add entry
entryCount = 0      'questions belonging to specific COI/MOE/MOP

Do Until rs.EOF

    If Trim(rs!MOP) = entry Then
        entryCount = entryCount + 1
    Else
        categoryCount.Add entryCount
        entryCount = 1
        entry = Trim(rs!MOP)
        categoryName.Add entry
    End If

    rs.MoveNext

    If rs.EOF Then
        categoryCount.Add entryCount
    End If

Loop

first = 1      'first record in recordset to put on chart

numChart = 0      'instantiate chart counter

'Step through collections generated in previous block, to determine position of first
'and last data records to send to Chart Wizard. Max chart size is set to 10 data
'points. Charts are evenly spaced down spreadsheet.

For i = 1 To categoryName.Count      'total number of unique categories
    num = categoryCount(i)      'number of questions pertinent to category i

    If (num > 10) Then

        For j = 0 To Int(num / 10)
            remainder = num - j * 10

            If remainder < 10 Then
                last = last + remainder 'last record to put on this chart
            Else
                last = last + 10      'last record to put on this chart
            End If

            chTitle = categoryName(i)
            makeChart first, last, openApps(k), numChart, chTitle 'produce chart
            If ((numChart) Mod (6)) = 0 Then      'condition for pagebreak
                If ((numChart) Mod (24)) = 0 Then      'every four pages
                    'open another instance of Excel due to chart limitations
                    Set oApp = CreateObject("Excel.Application")
                    newApp openApps, rs, oApp, k
                Else
                    'force a pagebreak
                    openApps(k).Worksheets("Sheet2").Rows(51 * _
                        ((numChart) Mod (24)) / 6).PageBreak = xlPageBreakManual
                End If
            End If
        Next j
    Else

```

```

last = last + num
chtTitle = categoryName(i)
makeChart first, last, openApps(k), numChart, chtTitle
If (((numChart) Mod (6)) = 0) Then
    If (((numChart) Mod (24)) = 0) Then
        Set oApp = CreateObject("Excel.Application")
        newApp openApps, rs, oApp, k
    Else
        openApps(k).Worksheets("Sheet2").Rows(51 * _
            ((numChart) Mod (24)) / 6).PageBreak = xlPageBreakManual
    End If
End If
End If

Next i

End If

SysCmd acSysCmdClearStatus

DoCmd.Hourglass False

For i = 1 To openApps.Count      'step through open Excel instances to view charts
    openApps(i).Visible = True
    openApps(i).Worksheets("Sheet2").PrintPreview
    If i < openApps.Count Then      'close all but the last
        openApps(i).DisplayAlerts = False
        openApps(i).Quit
    End If
Next i

Set oApp = Nothing

'AppActivate "Microsoft Access"
AppActivate "V-22 Questionnaire Tool (VQT)"

Exit_genGraphs_Click:
    Exit Sub

Err_genGraphs_Click:
    MsgBox Err.Description
    Resume Exit_genGraphs_Click

End Sub

Sub makeChart(first As Integer, last As Integer, oApp As Object, _
    numChart As Integer, chtTitle As String)
    'Create basic chartobject and chart using passed parameters to define data
    'used on chart and title of chart.

    Dim dataFirst As Variant, dataLast As Variant, labelFirst As Variant
    Dim labelLast As Variant, numDups As Variant
    Dim lr As Integer, ud As Integer, i As Integer
    Dim co As ChartObject
    Dim cht As Chart

    dataFirst = "D" & first      'create variants to define data range for this chart
    dataLast = "D" & last
    labelFirst = "B" & first
    labelLast = "B" & last

```

```

lr = (numChart) Mod (2)      'binary var to determine left/right chart placement
ud = Int(((numChart) Mod (24)) / 2)      'var to step charts down in pairs

'create chartobjects, place them, and size them according to # of datapoints

Set co = oApp.Worksheets(2).ChartObjects _
    .Add(50 + 225 * lr, ud * 217, 150, 100 + (last - first + 1) * 10)

numChart = numChart + 1

Set cht = co.Chart

oApp.Worksheets("Sheet1").Activate

buildIt oApp, dataFirst, dataLast, labelFirst, labelLast, chtTitle, cht

'for each datapoint in chart, pull # of responses from data table and add
'to chart as data label in place of data value

For i = 0 To (last - first)
    numDups = "C" & (first + i)
    If oApp.Worksheets("Sheet1").Range(numDups) > 0 Then
        With cht.SeriesCollection(1).Points(i + 1)
            .DataLabel.Text = oApp.Worksheets("Sheet1").Range(numDups)
        End With
    End If
Next i

first = last + 1

Set co = Nothing
Set cht = Nothing

End Sub

```

```

Sub buildIt(oApp As Object, dataFirst As Variant, dataLast As Variant, _
    labelFirst As Variant, labelLast As Variant, chtTitle As String, cht As Object)

'Create chart and format it according to passed parameters

With cht
    .ChartWizard _
        .Source:=oApp.Worksheets(1).Range(dataFirst, dataLast), _
        .Gallery:=xlBar, _
        .HasLegend:=False, _
        .Title:=chtTitle
    .SeriesCollection(1).XValues = oApp.Worksheets("Sheet1").Range(labelFirst, labelLast)
    .Axes(xlCategory).TickLabels.Font.Size = 6
End With

With cht.Axes(xlValue)
    .MinimumScale = 0
    .MaximumScale = 6
    .MinorUnit = 0.5
    .MajorUnit = 1
    .HasMajorGridlines = True
    .TickLabels.Font.Size = 6
End With

With cht.PlotArea.Interior
    .ColorIndex = 2
    .PatternColorIndex = 1

```

```

    .Pattern = xlSolid
End With

With cht.Axes(xlCategory, xlPrimary)
    .HasTitle = True
    .AxisTitle.Characters.Text = "Question"
End With

With cht.Axes(xlValue, xlPrimary)
    .HasTitle = True
    .AxisTitle.Characters.Text = "Average Response (# of responses)"
End With

With cht
    .Axes(xlValue).AxisTitle.Font.Size = 7
    .Axes(xlCategory).AxisTitle.Font.Size = 7
    .ApplyDataLabels =
        Type:=xlDataLabelsShowValue, _
        LegendKey:=False
End With

With cht.SeriesCollection(1).DataLabels
    .Font.Size = 8
    .NumberFormat = "0.00"
End With

End Sub

```

```

Sub newApp(openApps As Object, rs As Object, oApp As Object, k As Integer)
    'Add newly opened Excel application oApp to Collection openApps, add
    'workbook/sheets, and copy data from Access recordset rs to worksheet

    openApps.Add oApp

    k = k + 1

    openApps(k).UserControl = True
    openApps(k).Workbooks.Add

    rs.MoveFirst

    openApps(k).Worksheets("Sheet1").Range("A1").CopyFromRecordset rs

End Sub

```


LIST OF REFERENCES

1. "Corps Accepts First MV-22," *Marine Corps Gazette*, v. 83, no. 7, July 1999.
2. Schmidt, Nolan, V-22 Program Manager, "V-22 Program Brief for LtGen. Rhodes," NAS Patuxent River, MD, 1998.
3. Cwick, Mark, Senior Analyst, V-22 MOTT, Personal conversations with the author during November 1998, NAS Patuxent River, MD, 1998.
4. Naval Air Systems Command, PMA-275, *V-22 Test and Evaluation Master Plan (TEMP) No. M960*, Rev. B (draft), Washington, D.C., 1998.
5. Naval Air Systems Command, "U.S. Special Operations Command Requirements." [<http://www.navair.navy.mil/v22/ussocom.htm>]. 29 August 1999.
6. Department of Defense Systems Management College, *Test & Evaluation Management Guide*, 3rd ed., Ft. Belvoir, Defense Systems Management College Press, 1998.
7. Commander, Operational Test and Evaluation Force, U.S.N., *Test Plan for CNO Project M960 Operational Assessment/Evaluation (OT-IID, OT-IIE [OPEVAL] and Concurrent IT/OT Testing) of the V-22 Osprey*, Norfolk, 1998.
8. Stein, J. (Ed.), *The Random House College Dictionary, Revised Edition*, New York, Random House, Inc., 1982.
9. Cwick, Mark, Senior Analyst, V-22 MOTT, Multiple phone conversations with the author, November 1998 – July 1999.
10. McFedries, P., *Visual Basic for Applications Unleashed*, Sams Publishing, 1997.
11. Barker, F. S., *Access 97 Power Programming*, Que Corporation, 1997.
12. Roman, S., *Access Database Design & Programming*, O'Reilly & Associates, 1997.
13. Agresti, A., *Categorical Data Analysis*, New York, John Wiley & Sons, 1990.
14. Microsoft Corporation, "XL97: 'Not Enough Memory' Error Adding Chart to Workbook." [<http://www.microsoft.com/support/kb/articles/q168/6/50.asp>]. 02 July 1999.
15. Devore, J. L., *Probability and Statistics*, 4th ed., Pacific Grove, CA, Brooks/Cole Publishing Company, 1995.

16. Everitt, B.S., *Cluster Analysis*, 3rd ed., New York, Halsted Press, 1993.
17. Data Analysis Products Division, Mathsoft, *S-Plus 4 Guide to Statistics*, Mathsoft, 1997.

INITIAL DISTRIBUTION LIST

1. Defense Technical Information Center 2
8725 John J. Kingman Rd., STE 0944
Fort Belvoir, Virginia 22060-6218
2. Dudley Knox Library 2
Naval Postgraduate School
411 Dyer Rd.
Monterey, California 93943-5101
3. Director, Training and Education 1
MCCDC, Code C46
1019 Elliot Rd.
Quantico, Virginia 22134-5027
4. Director, Marine Corps Research Center 2
MCCDC, Code C40RC
2040 Broadway Street
Quantico, Virginia 22134-5107
5. Director, Studies and Analysis Division 1
MCCDC, Code C45
300 Russell Road
Quantico, Virginia 22134-5130
6. Marine Corps Representative 1
Naval Postgraduate School
Code 037, Bldg. 330, IN-116
555 Dyer Road
Monterey, California 93940
7. Marine Corps Tactical Systems Support Activity 1
Technical Advisory Branch
Attn: Maj J.C. Cummiskey
Box 555171
Camp Pendleton, California 92055-5080
8. Commanding Officer, V-22 MOTT 2
Attn: Operations Analyst
22474 James Road, Hangar 109
Unit ID 1 NAWCAD
Patuxent River, Maryland 20670-1624

9. Professor Lyn R. Whitaker, Code OR/Wh.....1
Department of Operations Research
Naval Postgraduate School
Monterey, California 93943-5000
10. Captain Thomas H. Hoivik, USN(Ret), Code OR/Ho1
Department of Operations Research
Naval Postgraduate School
Monterey, California 93943-5000
11. Professor Robert Read, Code OR/Re1
Department of Operations Research
Naval Postgraduate School
Monterey, California 93943-5000
12. Maj. William K. Hershberger.....2
61 Brittany Lane
Stafford, Virginia 22554